Dermoscopy, with and without visual inspection, for the diagnosis of melanoma in adults

Review information

Review type: Diagnostic test accuracy

Review number: #164

Authors

Jacqueline Dinnes[1], Jonathan J Deeks[1], Naomi Chuchu[1], Lavinia Ferrante di Ruffano[1], Rubeta N Matin[2], David R Thomson[3], Kai Yuen Wong[4], Roger Benjamin Aldridge[5], Rachel Abbott[6], Monica Fawzy[7], Susan E Bayliss[1], Matthew J Grainge[8], Yemisi Takwoingi[1], Clare Davenport[1], Kathie Godfrey[9], Fiona M Walter[10], Hywel C Williams[11], Cochrane Skin Cancer Diagnostic Test Accuracy Group[1]

[1]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[2]Department of Dermatology, Churchill Hospital, Oxford, UK
[3]Department of Plastic Surgery, St George's Hospital, London, UK
[4]Department of Plastic and Reconstructive Surgery, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[5]Department of Plastic Surgery, NHS Lothian/University of Edinburgh, Edinburgh, UK
[6]Welsh Institute of Dermatology, University Hospital of Wales, Cardiff, UK
[7]Department of Plastic and Reconstructive Surgery, Norfolk and Norwich University Hospital NHS Trust, Norwich, UK
[8]Division of Epidemiology and Public Health, School of Medicine, Nottingham, UK
[9]c/o Cochrane Skin Group, The University of Nottingham, Nottingham, UK
[10]Public Health & Primary Care, University of Cambridge, Cambridge, UK
[11]Centre of Evidence Based Dermatology, University of Nottingham, Nottingham, UK

Citation example: Dinnes J, Deeks JJ, Chuchu N, Ferrante di Ruffano L, Matin RN, Thomson DR, Wong KY, Aldridge RB, Abbott R, Fawzy M, Bayliss SE, Grainge MJ, Takwoingi Y, Davenport C, Godfrey K, Walter FM, Williams HC, Cochrane Skin Cancer Diagnostic Test Accuracy Group. Dermoscopy, with and without visual inspection, for the diagnosis of melanoma in adults. Cochrane Database of Systematic Reviews 2015 , Issue 10 . Art. No.: CD011902. DOI: 10.1002/14651858.CD011902 .

Contact person

Jacqueline Dinnes

Institute of Applied Health Research
University of Birmingham
Birmingham
B15 2TT
UK

E-mail: j.dinnes@bham.ac.uk

Dates

Assessed as Up-to-date: 29 August 2016
Date of Search: 29 August 2016
Next Stage Expected: 21 December 2017
Protocol First Published: Issue 10 , 2015
Review First Published: Not specified
Last Citation Issue: Issue 10 , 2015

What's new

Date	Event	Description

History

Date	Event	Description

Abstract

Background

Melanoma has one of the fastest rising incidence rates of any cancer. It accounts for a small percentage of skin cancer cases but is responsible for the majority of skin cancer deaths. Although history-taking and visual inspection of a suspicious lesion by a clinician are usually the first in a series of 'tests' to diagnose skin cancer, dermoscopy has become an important tool to assist diagnosis by specialist clinicians and is increasingly used in primary care settings. Dermoscopy is a magnification technique using visible light that allows more detailed examination of the skin compared to examination by the naked eye alone. Establishing the additive value of dermoscopy over and above visual inspection alone across a range of observers and settings is critical to understanding its contribution for the diagnosis of melanoma and to future understanding of the potential role of the growing number of other high-resolution image analysis techniques.

Objectives

To determine the diagnostic accuracy of dermoscopy for the detection of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants in adults, and to compare its accuracy with that of visual inspection alone. Studies were separated according to whether the diagnosis was recorded face-to-face (in-person) or based on remote (image-based) assessment.

Search methods

We undertook a comprehensive search of the following databases from inception up to August 2016: Cochrane Central Register of Controlled Trials; MEDLINE; Embase; CINAHL; CPCI; Zetoc; Science Citation Index; US National Institutes of Health Ongoing Trials Register; NIHR Clinical Research Network Portfolio Database; and the World Health Organization International Clinical Trials Registry Platform. We studied reference lists and published systematic review articles.

Selection criteria

Studies of any design that evaluated dermoscopy in adults with lesions suspicious for melanoma, compared with a reference standard of either histological confirmation or clinical follow-up. Data on the accuracy of visual inspection, to allow comparisons of tests, was included only if reported in the included studies of dermoscopy.

Data collection and analysis

Two review authors independently extracted all data using a standardised data extraction and quality assessment form (based on QUADAS-2). We contacted authors of included studies where information related to the target condition or diagnostic threshold were missing. We estimated accuracy using hierarchical summary ROC methods. Analysis of studies allowing direct comparison between tests was undertaken. To facilitate interpretation of results, we computed values of sensitivity at the point on the SROC curve with 80% fixed specificity and values of specificity with 80% fixed sensitivity. We investigated the impact of in-person test interpretation; use of a purposely developed algorithm to assist diagnosis; observer expertise; and dermoscopy training.

Main results

A total of 104 study publications reporting on 103 study cohorts with 42,788 lesions (including 5700 cases) were included, providing 354 datasets for dermoscopy. The risk of bias was mainly low for the index test and reference standard domains and mainly high or unclear for participant selection and participant flow. Concerns regarding the applicability of study findings were largely scored as 'High' concern in three of four domains assessed. Selective participant recruitment, lack of reproducibility of diagnostic thresholds and lack of detail on observer expertise were particularly problematic.

The accuracy of dermoscopy for the detection of invasive melanoma or atypical intraepidermal melanocytic variants was reported in 86 datasets; 26 for evaluations conducted in-person (dermoscopy added to visual inspection) and 60 for image-based evaluations (diagnosis based on interpretation of dermoscopic images). Analyses of studies by prior testing revealed no obvious effect on accuracy; analyses were hampered by the lack of studies in primary care, lack of relevant information and the restricted inclusion of lesions selected for biopsy or excision. Accuracy was higher for in-person diagnosis compared to image-based evaluations (relative diagnostic odds ratio (RDOR) of 4.6; 95% CI 2.4, 9.0, P<0.001).

Accuracy was compared for (a) in-person evaluations of dermoscopy (26 evaluations; 23,169 lesions and 1664 melanomas) versus visual inspection alone (13 evaluations; 6740 lesions and 459 melanomas) and for (b) image-based evaluations of dermoscopy (60 evaluations; 13,475 lesions and 2851 melanomas) versus image-based visual inspection (11

evaluations; 1740 lesions and 305 melanomas). For both comparisons, meta-analysis found dermoscopy to be more accurate than visual inspection alone, with RDORs of (a) 4.7 (95% CI: 3.0 to 7.5; P < 0.001) and (b) 5.6 (95% CI: 3.7 to 8.5; P < 0.001). These effects correspond to predicted differences in sensitivity of (a) 16% (95% CI: 8%, 23%) (92% for dermoscopy+visual inspection vs 76% for visual inspection) and (b) 35% (95% CI 24% to 46%) (81% for dermoscopy vs 47% for visual inspection) at a fixed specificity of 80%; and to predicted differences in specificity of (a) 20% (95% CI 7%, 33) (95% for dermoscopy plus visual inspection vs 75% for visual inspection) and (b) 40% (95% CI 27, 57) (82% for dermoscopy vs 42% for visual inspection) at a fixed sensitivity of 80%.

Using the median prevalence of disease in each set of studies ((a) 12% for in-person and (b) 24% for image-based) for a hypothetical population of 1000 lesions, an increase in sensitivity of (a) 16% (in-person) and (b) 35% (image-based) from using dermoscopy at a fixed specificity of 80% equates to a reduction in the number of melanomas missed of (a) 19 and (b) 81 with (a) 176 and (b) 152 false positive results. An increase in specificity of (a) 20% (in-person) and (b) 40% (image-based) at a fixed sensitivity of 80% equates to a reduction in the number of unnecessary excisions from using dermoscopy of (a) 176 and (b) 304 with (a) 24 and (b) 48 melanomas missed.

The use of a named or published algorithm to assist dermoscopy interpretation (as opposed to no reported algorithm or reported use of pattern analysis) had no significant impact on accuracy either for in-person (RDOR 1.4, 95% CI 0.34, 5.6; P=0.17) or image-based (RDOR 1.4, 95% CI 0.60, 3.3; P=0.22) evaluations. This result was supported by subgroup analysis according to algorithm used. Higher accuracy for observers reported as having high experience and for those classed as ‘expert consultants’ in comparison to those considered to have less experience in dermoscopy was observed, particularly for image-based evaluations. Evidence for the effect of dermoscopy training on test accuracy was very limited but suggested associated improvements in sensitivity.

Authors' conclusions

Despite the observed limitations in the evidence base, dermoscopy is a valuable tool to support the visual inspection of a suspicious skin lesion for the detection of melanoma and atypical intraepidermal melanocytic variants, particularly in referred populations and in the hands of experienced users. Data to support its use in primary care is limited however it may assist in triaging suspicious lesions for urgent referral when employed by suitably trained clinicians. Formal algorithms may be of most use for dermoscopy training purposes and for less expert observers, however reliable data comparing approaches using dermoscopy in-person are lacking.

Plain language summary

What is the diagnostic accuracy of dermoscopy in comparison to visual inspection of skin lesions for the diagnosis of melanoma in adults?

What is the aim of the review?

The aim of this Cochrane Review was to find out the accuracy of dermoscopy for the diagnosis of melanoma in comparison to visual inspection of suspicious skin lesions with the naked eye. The Review also investigated whether diagnostic accuracy using dermoscopy on a patient in-person differed to the accuracy of diagnosis using dermoscopic images of suspicious skin lesions. Researchers in Cochrane included 104 studies to answer this question.

Why is improving the diagnosis of melanoma important?

Melanoma is one of the most dangerous forms of skin cancer. Not recognising a melanoma when it is present (a false negative test result) delays surgery to remove it, risking cancer spreading to other organs in the body and possibly death. Diagnosing a skin lesion as a melanoma when it is not (a false positive result) may result in unnecessary surgery, further investigations and patient anxiety. Visual inspection of suspicious skin lesions by a clinician using the naked eye is usually the first of a series of ‘tests’ to diagnose melanoma. Magnification techniques can be used by skin cancer specialists to allow a more detailed examination of suspicious skin lesions than can be achieved using the naked eye alone.

What was studied in the review?

Dermoscopy is a handheld device using visible light (such as from incandescent or LED bulbs) that can be used as part of the clinical examination of suspicious skin lesions. Dermoscopy has become an important tool to assist diagnosis by specialist clinicians and is also increasingly used in primary care settings. Knowing the diagnostic accuracy of dermoscopy added to visual inspection alone is important to understanding who it should be used by and in which healthcare settings.

Researchers sought to find out the diagnostic accuracy of dermoscopy of suspicious skin lesions on a patient in-person and using dermoscopic images compared to visual inspection alone. Researchers also sought to find out whether diagnostic accuracy was improved by use of a dermoscopy checklist or by an increase in level of clinical expertise.

What are the main results of the review?

The review included 104 studies reporting data for people with lesions suspected of melanoma. The main results for the diagnosis of melanoma (including very early melanomas) are based on 86 of the studies, 26 of which provide information on the accuracy of dermoscopy added to in-person visual inspection of a skin lesion and 60 provide information based on examination of dermoscopic images without the patient being present.

The 26 in-person studies provide the most relevant data for the use of dermoscopy in practice and their results are summarised here. A total of 23,169 suspicious skin lesions were included in the 26 studies and 13 of them also provided information on the accuracy of visual inspection of a lesion without the use of dermoscopy. The results suggest that dermoscopy is more accurate than visual inspection on its own both for identifying melanoma correctly and excluding things that are not melanoma.

The studies used different ways of deciding whether a skin lesion was a melanoma or not which means that we cannot be exactly sure about how much better dermoscopy is compared to visual inspection alone. Instead we can give an illustrative example of the expected effect of the increase in accuracy using a group of 1000 lesions, of which 120 (12%) are melanoma. In order to see how much better dermoscopy is in identifying melanoma correctly when compared to just looking at the skin, we have to assume that both lead to the same number of lesions being falsely diagnosed as melanoma (we assumed that 176 of the 880 lesions without melanoma would have an incorrect diagnosis of melanoma). In this fixed situation, adding dermoscopy to visual inspection would correctly identify an extra 19 melanomas (110 compared with 91) that would have been missed by just looking at the skin alone. In other words, more melanomas would be correctly identified.

In order to see how much better dermoscopy is in deciding if a skin lesion is *not* a melanoma when compared to just looking at the skin, we have to assume that both lead to the same number of melanomas being correctly diagnosed (in this case we assumed that 96 out of the 120 melanomas would be correctly diagnosed). In this situation, adding in dermoscopy to visual inspection would reduce the number of lesions being wrongly diagnosed as being melanoma by 176 (a reduction from 220 in the visual inspection group to 44 lesions in the dermoscopy group). In other words, more lesions that were not melanoma would be correctly identified and less people would end up being sent for surgery.

Value of visual inspection checklists and effect of observer expertise

There was no evidence that use of a checklist to help dermoscopy interpretation changed diagnostic accuracy. Accuracy was better (with fewer missed melanomas and fewer people having unnecessary surgery) when the diagnosis was made by people with more clinical expertise and training.

How reliable are the results of the studies of this review?

In the majority of included studies, the diagnosis of melanoma was made by lesion biopsy and the absence of melanoma was confirmed by biopsy or by follow up over time to make sure the skin lesion remained negative for melanoma, both of which are likely to have been a reliable method for deciding whether patients really had melanoma. In a few studies, the absence of melanoma was made by expert diagnosis, which is unlikely to have been a reliable method for deciding whether patients really had melanoma. Poor reporting of study conduct made assessment of the reliability of studies difficult. Selective participant recruitment, lack of detail regarding the threshold for deciding on a positive test result were particularly problematic.

Who do the results of this review apply to?

Sixty-six studies were undertaken in Europe (77%), with the remainder undertaken in North America (n=6), Asia (n=4), Oceania (n=4), or were multicentre (n=7). Mean age ranged from 30 to 58 years (reported in 26 studies). The percentage of individuals with melanoma ranged between 1% and 41% for dermoscopy in-person studies (median 12%) and between 3% and 61% in studies using dermoscopy images (median 24%). Almost all of the studies were carried out in referral settings rather than in primary care. In the majority of studies the lesions were unlikely to be representative of the range of those seen in practice, for example only including skin lesions of a certain size or with a specific appearance. In addition variation in the expertise of clinicians performing visual inspection and the definition used for a positive dermoscopy test result across studies makes it unclear as to how dermoscopy should be carried out and by people with different levels of clinical expertise in order to achieve the accuracy observed in studies.

What are the implications of this review?

When used by specialists, dermoscopy is better at diagnosing melanoma compared to inspection of a suspicious skin lesion using the naked eye alone. Dermoscopy is more accurate when interpreted with the patient present rather than using dermoscopy images. Dermoscopy might help general practitioners to correctly identify people with suspicious lesions who need to be seen by a specialist. Checklists to help interpret dermoscopy might improve the accuracy of people with less expertise and training. Further, well reported studies assessing the diagnostic accuracy of dermoscopy when used in primary care and to identify the best way of delivering dermoscopy training are needed.

How up-to-date is this review?

The review authors searched for and used studies published up to August 2016.

*In these studies biopsy, clinical follow up or specialist clinician diagnosis were the reference standards.

Background

This review is one of a series of Cochrane Diagnostic Test Accuracy (DTA) reviews on the diagnosis and staging of melanoma and keratinocyte skin cancers conducted for the National Institute for Health Research (NIHR) Cochrane Systematic Reviews Programme. Appendix 1 shows the content and structure of the programme.

Target condition being diagnosed

Melanoma is one of the most aggressive forms of skin cancer, with the potential to metastasise to other parts of the body via the lymphatic system and blood stream. It accounts for a small percentage of skin cancer cases but is responsible for up to 75% of skin cancer deaths (Boring 1994; Cancer Research UK 2017a).

Melanoma arises from uncontrolled proliferation of melanocytes - the epidermal cells that produce pigment or melanin. It most commonly arises in the skin but can occur in any organ that contains melanocytes, including mucosal surfaces, the back of the eye, and lining around the spinal cord and brain. Cutaneous melanoma refers to a skin lesion with malignant melanocytes present in the dermis, and includes superficial spreading, nodular, acral lentiginous, and lentigo maligna melanoma variants (see Figure 1). Melanoma *in situ* refers to malignant melanocytes that are contained within the epidermis and have not yet invaded the dermis, but are at risk of progression to melanoma if left untreated. Lentigo maligna, a subtype of melanoma-in-situ in chronically sun-damaged skin, denotes another form of proliferation of abnormal melanocytes. Lentigo maligna can progress to invasive melanoma if its growth breaches the dermo-epidermal junction during a vertical growth phase (when it becomes known as 'lentigo maligna melanoma'), however its rate of malignant transformation is both lower and slower than for melanoma *in situ* (Kasprzak 2015). Melanoma *in situ* and lentigo maligna are both atypical intraepidermal melanocytic variants.

The incidence of melanoma rose to over 200,000 newly diagnosed cases worldwide in 2012 (Erdmann 2013; Ferlay 2015), with an estimated 55,000 deaths (Ferlay 2015). The highest incidence is observed in Australia with 11,405 new cases of melanoma of the skin (ACIM 2014) and in New Zealand with 2,341 registered cases (HPA and MelNet NZ 2014) in 2010. For 2014 in the USA, the predicted incidence was 73,870 per annum and the predicted number of deaths 9,940 (Siegel 2015). The highest rates in Europe are seen in north-western Europe and the Scandinavian countries, with highest incidence reported in Switzerland of 25.8 per 100,000 in 2012. Rates in the UK have trebled from 4.6 and 6.0 per 100,000 in men and women, respectively in England in 1990, to 18.6 and 19.6 per 100,000 in 2012 (EUCAN 2012). In the UK, melanoma has one of the fastest rising incidence rates of any cancer, and has the biggest projected increase in incidence between 2007 and 2030 (Mistry 2011). In the decade leading up to 2013, age standardised incidence increased by 46%, with 14,500 new cases in 2013 and 2,459 deaths in 2014 (Cancer Research UK 2017b). While overall incidence rates are higher in women than in men, the rate of incidence in the latter is increasing faster than in women (Arnold 2014).

The rising incidence in melanoma is thought to be primarily related to an increase in recreational sun exposure and tanning bed use and an increasingly ageing population with higher lifetime ultraviolet (UV) exposure, in conjunction with possible earlier detection (Linos 2009; Belbasis 2016). Putative risk factors are reviewed in detail elsewhere (Belbasis 2016), but can be broadly divided into host or environmental factors. Host factors include fair skin and light hair or eye colour; older age (Geller 2002); male sex (Geller 2002); previous skin cancer history (Tucker 1985); predisposing skin lesions, e.g., high melanocytic naevus counts (Gandini 2005), clinically atypical naevi (Gandini 2005), or large congenital naevi (Swerdlow 1995); genetically inherited skin disorders e.g., xeroderma pigmentosum (Lehmann 2011); and a family history of melanoma (Gandini 2005). Environmental factors include recreational and occupational exposure to sunlight (both cumulative and episodic burning) (Gandini 2005; Armstrong 2017); artificial tanning (Boniol 2012); and immunosuppression, e.g., in organ transplant recipients or human immunodeficiency virus (HIV)-positive individuals (DePry 2011). Lower socioeconomic class may be associated with delayed presentation and thus more advanced disease at diagnosis (Reyes-Ortiz 2006).

A database of over 40,000 US patients from 1998 onwards which assisted the development of the 8th American Joint Committee on Cancer (AJCC) Staging System indicated a five-year survival of 99% for very early stage melanoma, dropping to anything between 32% and 93% in stage III disease depending on tumour thickness, the presence of ulceration and number of involved nodes (Gershenwald 2017). Before the advent of targeted and immunotherapies, disseminated melanoma (to distant sites / visceral organs) was associated with median survival of six to nine months, one year survival rate of 25%, and three year survival of 15% (Balch 2009; Korn 2008).

Between 1975 and 2010, five year relative survival for melanoma (i.e. not including deaths from other causes) in the US increased from 80% to 94%, with survival for localised, regional, and distant disease estimated at 99%, 70%, and 18%, respectively in 2010 (Cho 2014). However, mortality rates showed little change, at 2.1 per 100,000 deaths in 1975 and 2.7 per 100,000 in 2010 (Cho 2014). Increasing incidence in localised disease over the same period (from 5.7 to 21 per 100,000) suggests that much of the observed improvement in survival may be due to earlier detection and heightened vigilance (Cho 2014). New targeted therapies for advanced (stage IV) melanoma (e.g. BRAF inhibitors) have improved survival and immunotherapies are evolving such that long term survival is being documented (Pasquali 2018; Rozeman 2017). No new data regarding the survival prospects for patients with stage IV disease were analysed for the AJCC 8 staging guidelines due to lack of contemporary data (Gershenwald 2017).

Treatment of melanoma

For primary melanoma, the mainstay of definitive treatment is early detection and excision of the lesion, to remove both the tumour and any malignant cells that might have spread into the surrounding skin (Sladden 2009; Marsden 2010; NICE 2015a; Garbe 2016; SIGN 2017). Recommended surgical margins vary according to tumour thickness (Garbe 2016) and stage of disease at presentation (NICE 2015a).

Index test(s)

For the purposes of our series of reviews, each component of the diagnostic process, including visual inspection or clinical examination, is considered a diagnostic or index 'test', the accuracy of which can be established in comparison with a reference standard of diagnosis, either alone or in combination with other available technologies that may assist the diagnostic process. In this review, although dermoscopy is the primary focus, two index tests are in fact under consideration, namely visual inspection and dermoscopy both of which can be undertaken in-person (face-to-face with the patient) or image-based (remote from the patient using images). As dermoscopy is added to visual inspection of a skin lesion when it is undertaken in-person, we effectively have three index tests: visual inspection alone (in-person or using images), visual inspection plus dermoscopy (in-person dermoscopy), and dermoscopy alone (image-based dermoscopy).

As visual inspection of a lesion is always undertaken first in a face-to-face patient-consultation, in this section we first consider visual inspection alone before going on to describe the addition of dermoscopy.

Visual inspection

Clinical history taking to identify risk factors and visual inspection of the lesion, surrounding skin and comparison with other lesions on the rest of the body, is fundamental to the diagnosis of skin cancer. In the UK, clinical examination is typically done at two decision points – first in the general practice (GP) surgery where a decision is made to refer or not to refer, and then a second time by a dermatologist or other secondary care clinician where a decision is made to biopsy or not.

Visual inspection of a lesion relies on both non-analytical and analytical pattern recognition strategies (Elstein 2002; Norman 1989; Norman 2009). Non-analytical pattern recognition formulates an initial hypothesis hidden from the conscious view of the diagnostician, while analytical pattern recognition uses more explicit rules based on conscious analytical reasoning (Norman 2009). The balance between non-analytical and analytical reasoning varies between clinicians, according to factors such as constitutional reasoning style preference, experience and familiarity with the diagnostic question. Various attempts have been made to formalise the "mental rules" involved in analytical pattern recognition for melanoma, ranging from setting out criteria that should be considered (e.g. 'pattern analysis'; Friedman 1985; Sober 1979) to formal scoring systems with explicit numerical thresholds (MacKie 1985; MacKie 1990). These variants on visual inspection strategies, and their comparative accuracy, are reviewed in detail in a separate systematic review in this series (Dinnes 2018a). Data on the accuracy of visual inspection has been included in this review only where both visual inspection and dermoscopy were evaluated in the same lesions in order to robustly estimate the comparative accuracy of adding dermoscopy to visual inspection compared to visual inspection alone, so that the benefit of dermoscopy can be quantified.

Visual inspection of a digital photograph (or 'macroscopic' image) of a suspicious skin lesion can also be undertaken as part of a teledermatology consultation whereby photographs, dermoscopic images, or both are taken by non-specialist clinicians and forwarded to a dermatologist, to obtain a specialist opinion (Chuchu 2018a). Images can also be encompassed in a store-and-forward smartphone application whereby a photograph of a concerning lesion is taken by the smartphone user and forwarded for an assessment of skin cancer risk by a specialist clinician (Chuchu 2018b). Images are often accompanied by a summary of the medical history and demographic information as part of a consultation package (Ndegwa 2010). According to UK guidelines, both clinical and dermoscopic images must be sent for 'full dermatology', i.e. as a replacement for a face-to-face consultation, whereas for 'triage teledermatology' dermoscopic images should be sent where facilities permit (BAD 2013).

Dermoscopy

Dermoscopy (also referred to as dermatoscopy or epiluminescence microscopy or ELM) has become a widely used tool for the specialist clinician and is increasingly being used in primary care settings. It uses a hand-held microscope and incident light (with or without oil immersion) to reveal subsurface images of the skin at increased magnification of x 10 to x 100 (Kittler 2011). Used alongside clinical examination, dermoscopy has been shown in some studies to increase the sensitivity of clinical diagnosis of melanoma from around 60% to as much as 90% (Kittler 1999; Carli 2002; Bono 2006; Stanganelli 2000) with much smaller effects in others (Benelli 1999; Bono 2002).

The visual nature of dermoscopic interpretation means that when used on an in-person basis, dermoscopy is essentially added to visual inspection of a skin lesion and similar non-analytical and analytical pattern recognition strategies are also employed to reach a diagnosis. Pattern analysis (Steiner 1987; Pehamberger 1993) is thought to be the most specific and reliable technique to aid dermoscopy interpretation when used by specialists (Maley 2014); however, dermoscopic histological correlations have been established and diagnostic algorithms developed based on colour, aspect, pigmentation pattern, and skin vessels. One of the first formal scoring systems was the ABCD rule for dermoscopy (Nachbar 1994; Stolz 1994), which includes 21 different features to be considered and scored (two based on asymmetry of the lesion, 8 on lesion border, 6 related to lesion colour and five to differential structures), and has reported sensitivity ranging between 84% and 93% (Nachbar 1994; Stolz 1994). Subsequently published algorithms attempt to simplify assessment without missing melanomas, for example, the Menzies tool (Menzies 1996), the seven-point dermoscopy checklist (Annessi 2007; Argenziano 1998; Argenziano 2001; Gereli 2010, amongst others), and the three-point checklist (Gereli 2010). However, dermoscopy can fail to diagnose atypical or early or featureless melanomas (Skvara 2005). These and other identified algorithms are described in detail in Appendix 2.

In modern practice, dermoscopic images are almost always obtained for skin lesions that are recommended for excision and are also obtained for lesions that have not yet met the diagnostic threshold for excision but are to be monitored over time in case of any further suspicious changes. Dermoscopic images are also a key component of teledermatology consultations, usually accompanied by digital photographs and other pertinent information (Chuchu 2018a), as discussed above.

The accuracy of dermoscopy has been suggested to vary with examiner experience (Kittler 2011), and results when used by untrained or less experienced examiners are potentially no better than clinical inspection alone (Binder 1997; Kittler 2002). Training in dermoscopy use can comprise from a single one-hour lecture (Benvenuto-Andrade 2006) to an intensive course lasting a week or more (De Giorgi 2011), often supplemented with web-based learning or using textbooks or CD-ROMs (Carli 2003; Menzies 2009; Tan 2009). The most effective means of training health professionals in dermoscopy remains to be established. Evidence from Australia suggests that it takes time to train non-expert clinicians in the use of dermoscopy, and dropout rates from training programmes may be up to 40% (Menzies 2009).

Clinical Pathway

The diagnosis of melanoma can take place in primary, secondary, and tertiary care settings by both generalist and specialist healthcare providers. In the UK, people with concerns about a new or changing lesion will usually present first to their general practitioner or less commonly directly to a specialist in secondary care, which could include a dermatologist, plastic surgeon, general surgeon or other specialist surgeon (such as an ear, nose, and throat (ENT) specialist or maxillofacial surgeon), or ophthalmologist (Figure 2). Current UK guidelines recommend that all suspicious pigmented lesions presenting in primary care should be assessed by taking a clinical history and visual inspection using the weighted seven-point checklist (MacKie 1990); lesions suspected to be melanoma should be referred for appropriate specialist assessment within two weeks (Chao 2013; Marsden 2010; NICE 2015a).

There are currently no recommendations promoting the use of dermoscopy in primary care in the UK, although the 2015 NICE suspected cancer recognition and referral guidelines states that people should be referred “using a suspected cancer pathway referral (for an appointment within 2 weeks) if dermoscopy suggests melanoma of the skin” (NICE 2015a). Studies from France (Chappuis 2016) and the Netherlands (Ahmadi 2017) suggest that around 8% of GPs use dermoscopy compared to as many as 40% of GPs in Australia reported to using a dermoscope in their routine practice (Youl 2007).

Theoretically, teledermatology consultations may aid appropriate triage of lesions into urgent referral; non-urgent secondary care referral (e.g., for suspected basal cell carcinoma); or where available, referral to an intermediate care setting, e.g. clinics run by GPs with a special interest in dermatology. The distinction between setting and examiner qualifications and experience is important as specialist clinicians might work in primary care settings (for example, in the UK, general practitioners (GPs) with a special interest in dermatology and skin surgery who have undergone appropriate training), and generalists might practice in secondary care settings (for example, plastic surgeons who do not specialise in skin cancer). The level of skill and experience in skin cancer diagnosis will vary for both generalist and specialist care providers and will also impact on test accuracy.

Following referral, a specialist clinician will also use history-taking and visual inspection of the lesion (in comparison with other lesions on the skin), usually in conjunction with dermoscopic examination, to inform a clinical decision. If melanoma is suspected, then urgent excision biopsy is recommended; for suspected cutaneous squamous cell carcinoma (cSCC) urgent excision with predetermined surgical margins. Other lesions such as basal cell carcinoma (BCC), suspected dysplastic naevi or pre-malignant lesions such as lentigo maligna may also be referred for a diagnostic biopsy, followed by appropriate treatment, further surveillance or reassurance and discharge.

Prior test(s)

Although smartphone applications and community-based teledermatology services can increasingly be directly accessed by people who have concerns about a skin lesion, visual inspection of a suspicious lesion by a clinician is usually the first in a series of tests to diagnose skin cancer. In the UK this usually takes place in primary care, however in some countries people with suspicious lesions can present directly to a specialist setting (NICE 2015b). Dermoscopy is likely to be added to visual inspection of a lesion in secondary care and referral settings, however, it is increasingly used in primary care, particularly in countries such as Australia (Youl 2007).

Consideration of the degree of prior testing that study participants have undergone is key to interpretation of test accuracy indices, as these are known to vary according to the disease spectrum (or case-mix) of included participants (Lachs 1992; Moons 1997; Leeflang 2013; Usher-Smith 2016). Spectrum effects are often observed when tests that are developed further down the referral pathway have lower sensitivity and higher specificity when applied in settings with participants with limited prior testing (Usher-Smith 2016). Studies of individuals with suspicious lesions at the initial clinical presentation stage ('test naïve') are likely to have a wider range of differential diagnoses and include a higher proportion of people with benign diagnoses compared with studies of participants who have been referred for a specialist opinion on the basis of visual inspection (with or without dermoscopy) by a generalist practitioner. Furthermore, studies in more specialist settings may focus on equivocal or difficult to diagnose lesions rather than lesions with a more general level of clinical suspicion. However this direction of effect is not consistent across tests and diseases, the mechanisms in action often being more complex than prevalence alone and can be difficult to identify (Leeflang 2013). A simple categorisation of studies according to primary, secondary or specialist setting therefore may not always adequately reflect this difference in disease spectrum.

Role of index test(s)

Although visual inspection and history-taking are key to diagnosing skin cancer and are always undertaken as part of a clinical examination, dermoscopy has become an important tool to assist diagnosis by specialist clinicians and is increasingly used in primary care settings. For the majority of generalist practitioners, the primary goal is to identify people with benign lesions and appropriately reassure them, thereby minimising the proportion of people who are referred unnecessarily, while still identifying those lesions that require referral and expert assessment. For the specialist, the aim is not only to identify those in need of urgent excision due to invasive cancer, but also to identify high risk lesions with considerable potential to progress to invasive disease, such as those with severe dysplasia or *in situ* disease e.g. lentigo maligna, for example.

When diagnosing potentially life-threatening conditions such as melanoma, the consequences of falsely reassuring a person that they *do not* have skin cancer can be serious and potentially fatal, as the resulting delay to diagnosis means that the window for successful early treatment may be missed. To minimise such false-negative diagnoses, a good diagnostic test will demonstrate high sensitivity and a high negative predictive value (NPV), where very few of those with a negative test result will actually have a melanoma. False positive test results from a test with poor specificity will result in the removal of many benign lesions. Unneccessary surgery is arguably less of an error than missing a potentially fatal melanoma, but is costly: false-positive diagnoses not only cause unnecessary scarring from the biopsy or excision procedure, but also increase patient anxiety whilst they await the definite histology results and increase healthcare costs as the number needed to remove to yield one melanoma diagnosis increases.

The additive value of dermoscopy over and above visual inspection alone is likely to vary with differences in setting, prior testing and selection of participants, and observer qualifications, experience and training. Furthermore dermoscopic images of lesions are increasingly taken by non-expert clinicians or by non clinicians, sometimes using mobile phone applications, and are forwarded to specialist clinics or to commercial organisations for interpretation, sometimes accompanied by a clinical image of the lesion and/or with varying amounts of patient information (such as age, gender, and location of the lesion). With skin cancer rates continuing to rise, the increasing availability of dermoscopy for generalist use, and with a growing number of other high-resolution image analysis techniques particularly for specialist use, it is important to understand the relative accuracy and appropriate place of available tests in the diagnostic pathway (whether as replacements for dermoscopy, or as add-on diagnostic tools).

Although the accuracy of image-based dermoscopy interpretation will be examined in this review, studies conducted specifically in a teledermatology context are the subject of a separate systematic review (Chuchu 2018a). Similarly, studies of mobile phone applications where the intended users are members of the general public rather than clinicians are the subject of further review (Chuchu 2018b).

Alternative test(s)

A number of other tests which may have a role in the diagnosis of melanoma in a specialist setting have been reviewed as part of our series of systematic reviews, including reflectance confocal microscopy (RCM) (Dinnes 2018b), optical coherence tomography (OCT) (Ferrante di Ruffano 2018a), and computer-aided diagnosis (CAD) techniques applied to various types of images including those generated by dermoscopy, diffuse reflectance spectrophotometry (DRS) and electrical impedance spectroscopy (EIS) (Ferrante di Ruffano 2018b), and high frequency ultrasound (Dinnes 2018c). Other tests reviewed include teledermatology (Chuchu 2018a) and mobile phone applications (Chuchu 2018b). Evidence permitting, the accuracy of available tests will be compared in an overview review, exploiting within-study comparisons of tests and allowing the analysis and comparison of commonly used diagnostic strategies where tests may be used singly or in combination.

We also considered and excluded a number of tests from this review such as tests used for monitoring people (e.g. total body photography of those with large numbers of typical or atypical naevi). We also did not assess histopathological confirmation following lesion excision because it is the established reference standard for melanoma diagnosis and will be one of the standards against which the index tests are evaluated in these reviews.

Rationale

This series of reviews of diagnostic tests used to assist clinical diagnosis in either clinical practice or in a research setting, aims to identify the most accurate approaches to diagnosis and provide clinical and policy decision-makers with the highest possible standard of evidence on which to base diagnostic and treatment decisions. With increasing rates of melanoma and a trend to adopt the use of dermoscopy and other high-resolution image analysis in primary care, the anxiety around missing early cases needs to be balanced against the risk of over referrals, to avoid sending too many people with benign lesions for a specialist opinion. It is questionable whether all skin cancers identified by sophisticated techniques contribute to morbidity and mortality or whether newer technologies run the risk of increasing false-positive diagnoses. The full impact of use of these technologies cannot be understood without an understanding of the accuracy of more established techniques such as dermoscopy, in comparison to visual inspection. It is also possible that widespread use of dermoscopy in primary care with inadequate training could result in harm from missed melanomas, particularly if used as a replacement for traditional history-taking and clinical examination of the entire skin. Many branches of medicine have noted the danger of such "gizmo idolatry" amongst doctors (Leff 2008). The trend towards remote interpretation of clinical images (whether macroscopic or dermoscopic images of lesions) and the use of remote technologies that do not involve clinicians without substantive evidence could further disrupt clinical pathways and healthcare payments as they may attract custom from the worried well, leaving an ever decreasing pool of qualified doctors to pick up any resulting problems.

There are a number of available systematic reviews in the field. Some are limited by now out-of-date search periods, for example searches in Rajpara 2009 were carried out up to 2007, and in Vestergaard 2008 up to 2008. Others are focused on specific clinical questions, for example, selected health care professionals (Corbo 2012 including only direct comparisons of the accuracy of primary care physicians versus dermatologists, and Loescher 2011 reviewing the skin cancer detection skills of advanced practice nurses) or settings (Herschorn 2012 including direct comparisons of visual inspection versus dermoscopy in primary care). More recently, Harrington and colleagues (Harrington 2017) published a systematic review of clinical prediction rules (or published algorithms) to assist the diagnosis of melanoma (both for clinical examination and for dermoscopy) and included studies published up to May 2015. This review did not consider whether diagnoses were made based on images or were conducted in-person, nor did it consider variations in the definition of the target condition, and furthermore no comparison was made for diagnosis with and without the use of an algorithm.

The critical question about the accuracy of dermoscopy in addition to visual inspection and the impact of examiner, prior patient testing, underlying risk status and the use of images for diagnosis needs to be answered before the potential contribution of other diagnostic tests can be set in context and appropriately placed in the diagnostic pathway.

This review follows a generic protocol which covers the full series of Cochrane DTA reviews for the diagnosis of melanoma (Dinnes 2015a). The Background and Methods sections of this review therefore use some text that was originally published in the protocol (Dinnes 2015a) and text that overlaps some of our other reviews (Dinnes 2018a; Dinnes 2018b). Appendix 3 provides a glossary of terms used.

Objectives

To determine the diagnostic accuracy of dermoscopy alone, or when added to visual inspection of a skin lesion, for the detection of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants in adults.

Accuracy was estimated separately according to the prior testing undergone by study participants comparing those with limited prior testing with those referred for further evaluation of a suspicious lesion. We originally aimed to estimate the effect on accuracy of diagnosis based on a face-to-face (in-person) encounter versus a remote (image-based) assessment as a secondary objective, however given the considerable difference in nature of an in-person consultation compared to the viewing of an image, accuracy was estimated separately for each approach to diagnosis. We therefore aimed to compare tests in the following way:

To estimate incremental accuracy for the diagnosis of invasive melanoma and atypical intraepidermal melanocytic variants in adults, a) from dermoscopy added to in-person visual inspection of a skin lesion, or b) from dermoscopic image-based assessment in comparison to visual inspection of a clinical photograph.

Secondary objectives

For the identification of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants:

i. To compare the accuracy of dermoscopy to visual inspection alone, where both tests have been evaluated in the same studies (direct test comparisons)

ii. To determine the diagnostic accuracy of individual algorithms used to assist dermoscopy; and

iii. To determine the effect of observer experience on diagnostic accuracy

iv. To determine the effect of dermoscopy training on diagnostic accuracy

For the alternative definitions of the target condition:

v. To determine the diagnostic accuracy of dermoscopy alone, or added to visual inspection of a skin lesion, for the detection of invasive melanoma only in adults, and to estimate incremental accuracy a) from dermoscopy added to in-person visual inspection of a skin lesion, or b) from dermoscopic image-based assessment in comparison to visual inspection of a clinical photograph.

vi. To determine the diagnostic accuracy of dermoscopy alone, or added to visual inspection of a skin lesion, for the detection of any skin cancer or skin lesion with a high risk of progression to melanoma in adults, and to estimate incremental accuracy a) from dermoscopy added to in-person visual inspection of a skin lesion, or b) from dermoscopic image-based assessment in comparison to visual inspection of a clinical photograph.

Investigation of sources of heterogeneity

We set out to address a range of potential sources of heterogeneity for investigation across our series of reviews, as outlined in our generic protocol (Dinnes 2015a) and described in Appendix 4, however our ability to investigate these was necessarily limited by the available data on each individual test reviewed.

The sources of heterogeneity that were investigated for dermoscopy were:

- prior testing: comparing those at initial presentation versus referred patients
- in-person versus image-based evaluations
- type of reference standard: histology alone versus histology plus clinical follow-up or other reference standard
- use of a diagnostic algorithm: no algorithm reported versus any named algorithm used
- lesion type: pigmented versus melanocytic lesions
- number of observers making the diagnosis: single observer versus consensus of two or more
- disease prevalence: 0 to 5%; >5 to 10%, >10 to 20%, >20%

Methods

Criteria for considering studies for this review

Types of studies

We included test accuracy studies that allow comparison of the result of the index test with that of a reference standard, including the following:

- studies where all participants receive a single index test and a reference standard;
- studies where all participants receive more than one index test(s) and reference standard;
- studies where participants are allocated (by any method) to receive different index tests or combinations of index tests and all receive a reference standard (between-person comparative studies (BPC));
- studies that recruit series' of participants unselected by true disease status (referred to as case series for the purposes of this review);
- diagnostic case-control studies that separately recruit diseased and non-diseased groups (see Rutjes 2005);
- both prospective and retrospective studies; and
- studies where previously acquired clinical or dermoscopic images were retrieved and prospectively interpreted for study purposes.

We excluded studies from which we could not extract 2x2 contingency data or if they included less than five melanoma cases or less than five benign lesions. The size threshold of five is arbitrary. However such small studies are unlikely to add precision to estimate of accuracy.

Participants

We included studies in adults with pigmented skin lesions or lesions suspicious for melanoma or those at high risk of developing melanoma, including those with a family history or previous history of melanoma skin cancer, atypical or dysplastic naevus syndrome, or genetic cancer syndromes.

We excluded studies that recruited only participants with malignant diagnoses and studies that compared test results in participants with malignancy compared with test results based on 'normal' skin as controls, due to the inherent bias in such comparisons (Rutjes 2006).

We excluded studies conducted in children or which clearly reported inclusion of more than 50% of participants aged 16 and under.

Index tests

Studies reporting accuracy data for dermoscopy, with diagnosis made either in-person (face-to-face diagnosis) or image-based (diagnosis based on dermoscopic images, remotely from the study participant) were eligible for inclusion. All established algorithms or checklists to assist diagnosis were included.

Studies developing new algorithms or methods of diagnosis (i.e., derivation studies) were **included** if they:

- used a separate independent 'test set' of participants or images to evaluate the new approach, or
- investigated lesion characteristics that had previously been suggested as associated with melanoma and the study reported accuracy based on the presence or absence of specific combinations of characteristics.

Studies were **excluded** if they:

- used a statistical model to produce a data driven equation, or algorithm based on multiple diagnostic features, with no separate test set.
- used cross-validation approaches such as 'leave-one-out' cross-validation (Efron 1983)
- evaluated the accuracy of the presence or absence of individual lesion characteristics or morphological features, with no overall diagnosis of malignancy
- reported accuracy data for ‘clinical diagnosis’ with no clear description as to whether the reported data related to visual inspection alone or included dermoscopy in all study participants
- were based on the experience of a skin cancer-specific clinic, where dermoscopy may or may not have been used on an individual patient basis.

Although primary care clinicians can have a specialist interest in skin cancer, for the purposes of this review we considered primary care physicians as generalist practitioners and dermatologists as specialists. Within each group, we extracted any reporting of special interest or accreditation in skin cancer.

Target conditions

The primary target condition was defined as the detection of:

- any form of invasive cutaneous melanoma , or atypical intraepidermal melanocytic variants (i.e. including melanoma *in situ*, or lentigo maligna, which has a risk of progression to invasive melanoma).

Two additional definitions of the target condition were considered in secondary analyses, namely the detection of:

- any form of invasive cutaneous melanoma alone
- any skin lesion requiring excision. This latter definition includes other forms of skin cancer, such as basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (cSCC), as well as melanoma *in situ*, lentigo maligna, and lesions with severe melanocytic dysplasia.

The diagnosis of the keratinocyte skin cancers, basal cell carcinoma, and squamous cell carcinoma as primary target conditions using visual inspection and/or dermoscopy are the subject of a separate review (Dinnes 2018d).

Reference standards

The ideal reference standard is histopathological diagnosis in all eligible lesions. A qualified pathologist or dermatopathologist should perform histopathology. Ideally, reporting should be standardised detailing a minimum dataset to include the histopathological features of melanoma to determine the American Joint Committee on Cancer (AJCC) Staging System (e.g. Slater 2014). We did not apply this as a necessary inclusion criterion, but extracted any pertinent information.

Partial verification (applying the reference test only to a subset of those undergoing the index test) was of concern given that lesion excision or biopsy are unlikely to be carried out for all benign-appearing lesions within a representative population sample. Therefore to reflect what happens in reality, we accepted clinical follow-up of benign-appearing lesions as an eligible reference standard, whilst recognising the risk of differential verification bias (as misclassification rates of histopathology and follow-up will differ).

Additional eligible reference standards included cancer registry follow-up and 'expert opinion' with no histology or clinical follow-up. Cancer registry follow-up is considered less desirable than active clinical follow-up, as follow-up is not carried out within the control of the study investigators. Furthermore, if participant-based analyses as opposed to lesion-based analyses are presented, it may be difficult to determine whether the detection of a malignant lesion during follow-up is the same lesion that originally tested negative on the index test.

All of the above were considered eligible reference standards with the following caveats:

- all study participants with a final diagnosis of the target disorder must have a histological diagnosis, either subsequent to the application of the index test or after a period of clinical follow-up, and

- at least 50% of all participants with benign lesions must have either a histological diagnosis or clinical follow-up to confirm benignity.

Search methods for identification of studies

Electronic searches

The Information Specialist (SB) carried out a comprehensive search for published and unpublished studies. A single large literature search was conducted to cover all topics in the programme grant (see Appendix 1 for a summary of reviews included in the programme grant). This allowed for the screening of search results for potentially relevant papers for all reviews at the same time. A search combining disease-related terms with terms related to the test names, using both text words and subject headings was formulated (Appendix 5). The search strategy was designed to capture studies evaluating tests for the diagnosis or staging of skin cancer. As the majority of records were related to the searches for tests for staging of disease, a filter using terms related to cancer staging and to accuracy indices was applied to the staging test search, to try to eliminate irrelevant studies, for example, those using imaging tests to assess treatment effectiveness. A sample of 300 records that would be missed by applying this filter was screened and the filter adjusted to include potentially relevant studies. When piloted on MEDLINE, inclusion of the filter for the staging tests reduced the overall numbers by around 6000. The final search strategy, incorporating the filter, was subsequently applied to all bibliographic databases as listed below. The final search result was cross-checked against the list of studies included in five systematic reviews; our search identified all but one of the studies, and this study is not indexed on MEDLINE. The Information Specialist devised the search strategy, with input from the Information Specialist from Cochrane Skin. No additional limits were used.

We searched the following bibliographic databases to 29 August 2016 for relevant published studies:

- MEDLINE via OVID (from 1946);
- MEDLINE In-Process & Other Non-Indexed Citations via OVID; and
- EMBASE via OVID (from 1980).

We searched the following bibliographic databases to 30 August 2016 for relevant published studies:

- the Cochrane Central Register of Controlled Trials (CENTRAL) Issue 7, 2016, in the Cochrane Library;
- the Cochrane Database of Systematic Reviews (CDSR) Issue 8, 2016 in the Cochrane Library;
- Cochrane Database of Abstracts of Reviews of Effects (DARE) Issue 2, 2015;
- CRD HTA (Health Technology Assessment) database Issue 3, 2016;
- CINAHL (Cumulative Index to Nursing and Allied Health Literature via EBSCO from 1960).

We searched the following databases for relevant unpublished studies:

- CPCI (Conference Proceedings Citation Index) via Web of Science™ (from 1990);
- Zetoc (from 1993)
- SCI Science Citation Index Expanded™ via Web of Science™ (from 1900, using the "Proceedings and Meetings Abstracts" Limit function).

We searched the following trials registers:

- The US National Institutes of Health Ongoing Trials Register (www.clinicaltrials.gov);
- NIHR Clinical Research Network Portfolio Database (http://www.nihr.ac.uk/research-and-impact/nihr-clinical-research-network-portfolio/);
- The World Health Organization International Clinical Trials Registry Platform (apps.who.int/trialsearch/).

We aimed to identify all relevant studies regardless of language or publication status (published, unpublished, in press, or in progress). No date limits were applied.

Searching other resources

We have included information about potentially relevant ongoing studies in the 'Characteristics of ongoing studies' tables. We have screened relevant systematic reviews identified by the searches for their included primary studies, and included any missed by our searches. We have checked the reference lists of all included papers, and subject experts within the author team have reviewed the final list of included studies. No citation searching has been conducted.

Data collection and analysis

Selection of studies

Titles and abstracts were screened by at least one author (JDi or NC), with any queries discussed and resolved by consensus. A pilot screen of 539 MEDLINE references showed good agreement (89% with a kappa of 0.77) between screeners. Primary test accuracy studies and test accuracy reviews (for scanning of reference lists) of any test used to investigate suspected melanoma, BCC, or cSCC were included at initial screening. Inclusion criteria (Appendix 6) were applied independently by both a clinical reviewer (from one of a team of twelve clinician reviewers) and a methodologist reviewer (JDi or NC) to all full text articles, disagreements were resolved by consensus or by a third party (JDe, CD, HW, and RM). Authors of eligible studies were contacted when insufficient data were presented to allow for the construction of 2x2 contingency tables.

Data extraction and management

One clinical (as detailed above) and one methodologist reviewer (JDi, NC or LFR) independently extracted data concerning details of the study design, participants, index test(s) or test combinations and criteria for index test positivity, reference standards, and data required to populate a 2x2 diagnostic contingency table for each index test using a piloted data extraction form. Data were extracted at all available index test thresholds. Disagreements were resolved by consensus or by a third party (JDe, CD, HW, and RM).

Authors of included studies were contacted where information related to final lesion diagnoses or diagnostic threshold were missing. In particular, invasive cSCC (included as disease positive for one of our secondary objectives) is not always differentiated from '*in situ'* variants such as Bowens disease (which we did not consider as disease positive for any of our definitions of the target condition). Authors of conference abstracts published from 2013 to 2015 were contacted to ask whether full data were available. If no full paper was identified, we marked conference abstracts as 'pending' and will revisit them in a future review update.

Dealing with multiple publications and companion papers

Where multiple reports of a primary study were identified, we maximised yield of information by collating all available data. Where there were inconsistencies in reporting or overlapping study populations, we contacted study authors for clarification in the first instance. If this contact with authors was unsuccessful, we used the most complete and up-to-date data source where possible.

Assessment of methodological quality

We assessed risk of bias and applicability of included studies using the QUADAS-2 checklist (Whiting 2011), tailored to the review topic (see Appendix 7). The modified QUADAS-2 tool was piloted on a small number of included full text articles. One clinical (as detailed above) and one methodologist reviewer (JDi, NC or LFR) independently assessed quality for the remaining studies; any disagreement was resolved by consensus or by a third party where necessary (JDe, CD, HW, and RM).

Statistical analysis and data synthesis

Separate analyses were planned according to the point that study participants have reached in the clinical pathway (numbered from 1 to 7 in Figure 3), the clarity with which the pathway could be determined (clear or unclear), and the evaluation of in-person versus image-based diagnosis.

Our unit of analysis was the lesion rather than the patient. This is because (i) in skin cancer initial treatment is directed to the lesion rather than systemically (thus it is important to be able to correctly identify cancerous lesions for each person), and (ii) it is the most common way in which the primary studies reported data. Although there is a theoretical possibility of correlations of test errors when the same people contribute data for multiple lesions , most studies include very few people with multiple lesions and any potential impact on findings is likely to be very small, particularly in comparison with other concerns regarding risk of bias and applicability. For each analysis, only one dataset was included per study to avoid multiple counting of lesions. Where multiple algorithms were assessed in an individual study, datasets were selected on the following preferential basis:

i. 'no algorithm' reported; data presented for clinician's overall diagnosis or management decision

ii. pattern analysis or pattern recognition

iii. ABCD algorithm (or derivatives of)

iv. 7-point checklist (7PCL; also referred to as Glasgow/Mackie checklist)

v. Menzies algorithm

vi. 3-point checklist (3PCL)

For each index test, algorithm or checklist under consideration, estimates of sensitivity and specificity were plotted on coupled forest plots and in receiver operating characteristic (ROC) space. For tests where commonly used thresholds were reported we estimated summary operating points (summary sensitivities and specificities) with 95% confidence and prediction regions using the bivariate hierarchical model (Chu 2006; Reitsma 2005). Where inadequate data were available for the model to converge the model was simplified, first by assuming no correlation between estimates of sensitivity and specificity and secondly by setting estimates of near zero variance terms to zero (Takwoingi 2015). Where all studies reported 100% sensitivity (or 100% specificity) the number with disease (or no disease) was summed across studies and used to compute a binomial exact 95% confidence interval. Where reported, missing or indeterminate results were usually excluded by study authors such that data could not be included in our analyses. Where study authors reported missing or indeterminate results in more detail, these results were excluded by us for consistency.

Data on the accuracy of visual inspection, to allow comparisons of tests, was included only if reported in the studies of dermoscopy due to the known substantial unexplained heterogeneity in all studies of the accuracy of visual inspection (Dinnes 2018a). Comparisons were made between visual inspection results with dermoscopy data from all dermoscopy studies, and then only using dermoscopy data from studies that also reported visual inspection data for the same patients to enable a robust direct comparison (Takwoingi 2013).

We made comparisons between tests by comparing summary ROC curves using the hierarchical summary receiver-operator curves (HSROC) model (Rutter 2001) rather than by estimating average operating points as this approach allows incorporation of data at different thresholds and from different algorithms or checklists. We used an HSROC model that assumed a constant SROC shape between tests and subgroups (allowing for asymmetry in shape) and modelled differences in threshold and accuracy by addition of covariates. The significance of the differences between tests was assessed by the likelihood ratio test (LR test) assessing differences in both accuracy and threshold. Simpler models were fitted when convergence was not achieved due to small numbers of studies, first assuming symmetric SROC curves (setting the shape term to zero), and then setting random effects variance estimates to zero.

Estimates of accuracy from HSROC models are presented as diagnostic odds ratios (estimated where the SROC curve crosses the sensitivity=specificity line) with 95% confidence intervals. Differences between tests and subgroups from HSROC analyses are presented as relative diagnostic odds ratios with 95% confidence intervals. To facilitate interpretation in terms of rates of false positive and false negative diagnoses, values of sensitivity at the point on the SROC curve with 80% specificity and of specificity at the point on the SROC curve with 80% sensitivity have been computed. These 80% values were chosen as they lie within the estimates for the majority of analyses. These results should only be considered as illustrative examples of possible sensitivities (and specificities) and differences in sensitivities (and specificities) that could be expected. Confidence intervals for these estimates of sensitivity and specificity were computed assuming normal distribution of sampling error on logit scales; confidence intervals for differences in sensitivity and specificity were computed assuming normal distributions of sampling error on untransformed scales.

For computation of likely numbers of true positive, false positive, false negative and true negative findings in the summary of findings tables these indicative values were applied to lower quartile, median and upper quartiles of the prevalence observed in the study groups.

Bivariate models were fitted using the xtmelogit command in STATA 15 and HSROC models fitted using the NLMIXED procedure in the SAS statistical software package (SAS 2012, version 9.3; SAS Institute, Cary, NC, USA) and the metadas macro (Takwoingi 2010).

Investigations of heterogeneity

Investigations of heterogeneity, comparisons between algorithms and according to observer experience and qualifications were also made by comparing summary ROC curves using the hierarchical summary receiver-operator curves (HSROC) model (Rutter 2001) with additional covariates for differences in threshold and accuracy as used for comparing tests. Small subgroups were omitted from models where parameter estimates could not be obtained due to convergence problems.

Sensitivity analyses

Sensitivity analyses were planned, restricting analyses to studies where

- both dermoscopy (added to visual inspection) and visual inspection alone were evaluated in the same study (direct test comparisons as discussed above)
- partial verification was avoided (restricting to studies including follow-up of benign lesions)
- for studies using follow-up of benign appearing lesions, the interval between the index test and the reference standard was at least 3 months
- for direct test comparisons, the period of application between the index tests was within one month
- concerns around applicability for participant selection are low
- low risk of bias for the index test
- low risk of bias for the reference standard

Assessment of reporting bias

Because of uncertainty about the determinants of publication bias for diagnostic accuracy studies and the inadequacy of tests for detecting funnel plot asymmetry (Deeks 2005), we did not perform tests to detect publication bias.

Results

Results of the search

A total of 34,347 unique references were identified and screened for inclusion. Of these, 1051 full text papers were reviewed for eligibility for any one of the suite of reviews of tests to assist in the diagnosis of melanoma or keratinocyte skin cancer. Of the 1051 full text papers assessed, 848 were excluded from all reviews in our series (see Figure 4 PRISMA flow diagram of search and eligibility results).

Of the 340 studies tagged as potentially eligible for this review of dermoscopy, 104 publications were included. Exclusions were mainly due to the inability to construct a 2x2 contingency table based on the data presented (n=43); the use of ineligible index tests (n=19) (for example: reporting of data for 'clinical diagnosis' or for serial use of the index test in a follow-up context); assessment of individual lesion characteristics (n=28); or derivation type studies developing new algorithms or checklists without a separate training and test set of lesions (n=30). Other reasons for exclusion included not meeting our requirements for an eligible reference standard (n=15), ineligible study populations (n=21) (for example, recruiting only malignant or only benign lesions), inadequate sample size (n=22), ineligible definition of the target condition (n=18) or with test interpretation by medical students or laypersons (n=3). A list of the 236 publications excluded from this review with reasons for exclusion is provided in Characteristics of excluded studies, with a list of all studies excluded from the full series of reviews available as a separate pdf.

The authors of 17 publications were contacted for further data to allow study inclusion in the review; responses were received from four authors with regard to seven publications. Two authors provided additional data but this was insufficient to allow inclusion of the studies (Cabrijan 2008; Warshaw 2009; Warshaw 2009a; Warshaw 2010), one replied indicating that dermoscopy was not necessarily used in all study participants (Youl 2007; Youl 2007a) and one replied but was unable to access the data needed (Fabbrocini 2008). The authors of a further 20 included studies were contacted for further details of study methods. Responses were received in regard to 10 studies, 8 providing further information regarding the diagnostic thresholds used (Blum 2003; Blum 2004; Bono 2006; Bourne 2012; Carrera 2016; Durdu 2011; Kittler 1999; Stanganelli 2000), one providing full anonymised study data (Rosendahl 2011) and one unable to provide the information requested, although the study could still be included (Menzies 2009).

Of the 104 included study publications, two provide data for two separate cohorts of lesions: Guitera 2009 reports data for one cohort of lesions recruited in Modena, Italy (denoted Guitera 2009a (Modena)) and one cohort recruited in Sydney (denoted Guitera 2009b (Sydney)); Haenssle 2010 reports data for one cohort of lesions examined on participants first visit (denoted Haenssle 2010a (FV)) and one cohort of lesions identified during patient follow-up (denoted Haenssle 2010b (FU)). One further cohort is reported on in four different publications; data from one publication (Blum 2004b) have been included in the primary analyses with data from Blum 2003, Blum 2003b and Blum 2004 providing results for different algorithms or thresholds for the same set of lesions. The total number of cohorts of lesions described in the 104 study publications is therefore 103 (104 plus 2 minus 3). The 104 study publications provided a total of 354 dermoscopy datasets (each publication often providing more than one 2x2 contingency table according to the use of different algorithms, different test thresholds or different observers) for 42,788 lesions and 5700 malignancies. The total number of study participants with suspicious lesions cannot be estimated due to lack of reporting in study publications (reported in only 44 studies with 9591 participants). A third of study publications (n=31; 30%) also reported accuracy data for diagnosis using visual inspection; these provided 61 datasets for 9025 lesions and 959 malignancies. A systematic review of the accuracy of visual inspection *per se* is reported in Dinnes 2018a. A further 29 of the 104 included study publications reported data for tests other than dermoscopy or visual inspection including: teledermatology (n=3), reflectance confocal microscopy (RCM) (n=7), exfoliative cytology (n=1) and computer-assisted diagnosis (CAD) techniques (n=18).

Methodological quality of included studies

The overall methodological quality of all included studies (regardless of target condition) is summarized according to in-person or image-based approaches to dermoscopy or to visual inspection. A total of 35 in-person evaluations are presented in Figure 5 with results per study presented in Figure 6. A total of 74 image-based evaluations of dermoscopy are presented in Figure 7 with results per study presented in Figure 8. The total number of entries in Figure 6 and Figure 8 sums to 109 (35 + 74) instead 103 (as per the number of included cohorts) for the following reasons: a) three publications (Carli 2002; Dummer 1993; Unlu 2014) reported both in-person and image-based data and therefore appear in both the in-person and image-based plots (making 106 entries) and b) one cohort was reported on in four papers (Blum 2003; Blum 2003b; Blum 2004; Blum 2004b) which have all contributed data to the review analyses and therefore have been quality assessed four times (making 109 entries).

In-person evaluations

Risk of bias was judged to be 'Low' for the majority of studies in only two of five quality domains assessed (dermoscopy index test, reference standard); the majority of studies were at High or Unclear risk of bias for the remaining three domains (participant selection, visual inspection index test flow and timing) (Figure 5). Applicability of study findings were scored as of 'High' or 'Unclear' concern in all four domains (participant selection, dermoscopy and visual inspection index tests, reference standards) assessed.

For participant selection, 11 studies (31%) were judged at low risk of bias (Carli 1994; Dreiseitl 2009; Duff 2001; Grimaldi 2009; Langley 2007; Menzies 2009; Morales Callaghan 2008; Nachbar 1994; Soyer 2004; Stanganelli 2000; Unlu 2014); and 8 (23%) were considered high risk (Figure 6) due to exclusion of lesions by size (Bono 2002b; Bono 2006; Kittler 1999) or type (Ahnlide 2016; Cristofolini 1994; Guitera 2009a (Modena); Haenssle 2010a (FV); Haenssle 2010b (FU)). The study by Haenssle and colleagues excluded participants showing melanoma development on pre-existing pigmented lesions during the following 12 months after the analysed time frame. Twelve studies (34%) did not report the method of participant selection and 15 (43%) did not clearly describe exclusions from the study. Almost all cohorts (91%; n = 32) were considered at high concern for applicability of participants. In the majority of cases (n = 30), this was due to restricted study populations such as inclusion of only melanocytic (n = 10) lesions or inclusion of lesions selected for excision based on the clinical or dermoscopic diagnosis (n = 28). Only four cohorts (11%) were judged to have included a representative patient population (Dreiseitl 2009; Grimaldi 2009; Menzies 2009; Stanganelli 2000). Eight cohorts (23%) also included multiple lesions per participant (Durdu 2011; Gokdemir 2011; Grimaldi 2009; Haenssle 2010b (FU); Haenssle 2010a (FV); Kittler 1999; Morales Callaghan 2008; Stanganelli 2000) and 12 others (34%) did not clearly report number of included participants.

For the index test domain, there are 33 evaluations of in-person dermoscopy and 16 evaluations of in-person visual inspection (Figure 5). For dermoscopy, 24 evaluations (73%) were considered at low risk of bias (Ahnlide 2016; Argenziano 2006; Ascierto 2010; Bauer 2000; Benelli 1999; Bono 2002; Bono 2002b; Bono 2006; Broganelli 2005; Carli 1994; Carli 2002; Coras 2003; Cristofolini 1994; Durdu 2011; Feldmann 1998; Grimaldi 2009; Guitera 2009a (Modena); Haenssle 2010a (FV); Haenssle 2010b (FU); Langley 2007; Menzies 2009; Morales Callaghan 2008; Stanganelli 2000; Viglizzo 2004 and two evaluations (6%) were judged high risk (Kittler 1999; Nachbar 1994); 7 (21%) did not provide sufficient information to allow the risk of bias to be fully judged. All studies were judged to have made the diagnosis blinded to the reference standard result given that this is always undertaken prior to histology; 25 (76%) also clearly reported pre-specification of the diagnostic threshold - 20 using named algorithms or pattern analysis (Ahnlide 2016; Argenziano 2006; Ascierto 2010; Benelli 1999; Bono 2006; Broganelli 2005; Carli 1994; Carli 2002; Coras 2003; Cristofolini 1994; Durdu 2011; Feldmann 1998; Grimaldi 2009; Guitera 2009a (Modena); Haenssle 2010a (FV); Haenssle 2010b (FU); Langley 2007; Morales Callaghan 2008; Soyer 1995; Stanganelli 2000) and five (15%) (Bauer 2000; Bono 2002; Bono 2002b; Menzies 2009; Viglizzo 2004) describing the process by which the diagnosis was reached. Two studies (6%) developed new algorithms (Nachbar 1994) or evaluated multiple thresholds for test positivity (Kittler 1999).

All 16 visual inspection evaluations were also considered to have made the diagnosis blinded to the reference standard result. One (6%) was at high risk of bias due to evaluation of several different ABCDE algorithm thresholds (Benelli 1999) and 9 (56%) were judged unclear as to the diagnostic thresholds used.

High concern for the applicability of the index tests was recorded for 16 in-person evaluations of dermoscopy (48%) (Figure 5), primarily due to a lack of description of the diagnostic thresholds used (n=8, but also as a result of presentation of average (Argenziano 2006) or consensus diagnoses (Bauer 2000; Benelli 1999; Carli 1994; Carli 2002; Haenssle 2010b (FU); Haenssle 2010a (FV); Morales Callaghan 2008) as opposed to the diagnosis of a single observer. Six studies (18%) did not provide sufficient information to judge the clinical applicability of the dermoscopy diagnosis and observer expertise in dermoscopy could not be fully judged in five evaluations.

High concern for the applicability of the index tests was recorded for 14 of the 16 (88%) visual inspection evaluations Argenziano 2006; Benelli 1999; Bono 2002; Bono 2002b; Bono 2006; Carli 2002; Dummer 1993; Grimaldi 2009; Krahn 1998; Menzies 2009; Morales Callaghan 2008; Soyer 1995; Unlu 2014; Viglizzo 2004(, due to the threshold for diagnosis not detailed in 12 (75%), reporting of average (Argenziano 2006) or consensus (Benelli 1999; Carli 2002; Morales Callaghan 2008) diagnoses, or diagnosis by non-expert observers (Menzies 2009; Grimaldi 2009).

Of the 35 included in-person evaluations, 29 were judged at low risk of bias for the reference standard due to the use of an acceptable reference standard (83%) (Ahnlide 2016; Argenziano 2006; Ascierto 2010; Bauer 2000; Benelli 1999; Bono 2002; Bono 2002b; Bono 2006; Broganelli 2005; Carli 1994; Carli 2002; Coras 2003; Cristofolini 1994; Duff 2001; Dummer 1993; Durdu 2011; Feldmann 1998; Gokdemir 2011; Guitera 2009a (Modena); Kittler 1999; Krahn 1998; Langley 2007; Morales Callaghan 2008; Nachbar 1994; Piccolo 2000; Soyer 1995; Soyer 2004; Unlu 2014; Viglizzo 2004) (Figure 5). Five (14%) did not meet our criteria for an acceptable reference standard, with more than 20% of the benign lesions undergoing follow-up rather than excision (Grimaldi 2009; Haenssle 2010b (FU); Haenssle 2010a (FV); Menzies 2009; Stanganelli 2000) and one study was judged at unclear risk of bias due to lack of reporting of the number of patients with a histological reference standard and number with follow-up (Dreiseitl 2009). Blinding of the reference standard to the index test (in this case the pathology referral diagnosis) was recorded but did not contribute to the overall risk of bias for this domain. No blinding of the reference standard was implemented in Menzies 2009 and blinding was not described in 34 studies (97%). The applicability of the reference standard was of low concern in 7 evaluations (20%) (Argenziano 2006; Duff 2001; Feldmann 1998; Krahn 1998; Langley 2007; Nachbar 1994; Unlu 2014), high in one (Menzies 2009)and unclear for 27 (77%). In Menzies 2009, high concern was due to the use of expert opinion for classifying the final diagnosis of some lesions. Only 7 studies reported histopathology interpretation by an experienced histopathologist or by a dermatopathologist (Argenziano 2006; Duff 2001; Feldmann 1998; Krahn 1998; Langley 2007; Nachbar 1994; Unlu 2014).

In terms of flow and timing, 15 of the 35 cohorts were judged at high risk of bias (43%) (Ahnlide 2016; Argenziano 2006; Coras 2003; Dreiseitl 2009; Dummer 1993; Durdu 2011; Feldmann 1998; Grimaldi 2009; Guitera 2009a (Modena); Haenssle 2010a (FV); Haenssle 2010b (FU); Kittler 1999; Langley 2007; Menzies 2009; Stanganelli 2000), 6 (17%) at low risk (Ascierto 2010; Bauer 2000; Benelli 1999; Carli 1994; Morales Callaghan 2008; Soyer 2004) and 13 (37%) did not provide enough information on which to judge this domain (Figure 5). Of those at high risk, 6 evaluations did not use the same reference standard for all participants (differential verification) (Dreiseitl 2009; Grimaldi 2009; Haenssle 2010a (FV); Haenssle 2010b (FU); Menzies 2009; Stanganelli 2000), and 10 did not include all participants in the analysis (Ahnlide 2016; Argenziano 2006; Coras 2003; Dreiseitl 2009; Dummer 1993; Feldmann 1998; Guitera 2009a (Modena); Kittler 1999; Langley 2007; Menzies 2009). A further 23 (66%) cohorts were unclear on the interval between the application of the index test and excision for histology with only 6 (34%) reporting consecutive diagnosis and excision or biopsy (Ahnlide 2016; Ascierto 2010; Benelli 1999; Carli 1994; Durdu 2011; Feldmann 1998; Guitera 2009a (Modena); Haenssle 2010a (FV); Haenssle 2010b (FU); Langley 2007; Morales Callaghan 2008; Soyer 2004).

Image-based evaluations

Across the 74 image-based dermoscopy evaluations, risk of bias was judged to at High or Unclear in all domains apart from the dermoscopy index test domain (Figure 7 and Figure 8). Applicability of study findings were scored as of 'High' concern in almost all studies for three out of four domains assessed, only the reference standard domain raised few concerns about applicability.

For participant selection, 38 of the 74 evaluations (51%) were judged at high risk of bias (Arevalo 2008; Argenziano 2011; Benelli 2000; Binder 1994; Blum 2003b; Blum 2004; Carli 2003b; Carrera 2016; di Meo 2016; Feci 2015; Ferrari 2015; Friedman 2008; Gereli 2010; Guitera 2009b (Sydney); Hauschild 2014; Kittler 2001; Malvehy 2014; Menzies 2005; Menzies 2008; Menzies 2013; Piccolo 2014; Pizzichetta 2002; Pizzichetta 2004; Pupelli 2013; Rosendahl 2011; Rubegni 2012; Sboner 2004; Seidenari 1998; Skvara 2005; Stanganelli 1998; Stanganelli 1999; Stolz 1994; Tan 2009; Tenenhaus 2010; Troyanova 2003; Wells 2012; Westerhoff 2000; Winkelmann 2016) and 27 (36%) did not provide sufficient information to judge this domain (Figure 7). Nineteen evaluations (26%) implemented a case-control type design with separate sampling of melanoma and non melanoma lesions, and 25 (34%) excluded lesions: on the basis of size or thickness (n = 6); type of lesion (n = 8); lesion site (n = 3); equivocal pathology (n = 4); or inadequate image quality (n = 8). Twenty-nine evaluations (39%) did not report the method of participant selection and 31 (42%) did not clearly describe exclusions from the study. All evaluation cohorts were considered at high concern for applicability of participants. In the majority of cases, this was due to restricted study populations such as inclusion of only melanocytic (n = 35), amelanotic (n = 2), nodular (n = 1), regressing (n = 1) or acral (n = 1) lesions, or inclusion of lesions selected for excision based on the clinical or dermoscopic diagnosis (n = 57). Nineteen evaluations clearly reported including similar numbers of participants and lesions, seven reported inclusion of multiple lesions per participant and 48 did not report the number of participants.

For the index test domain, there are 74 evaluations of image-based dermoscopy and 15 evaluations of visual inspection of clinical images (Figure 7). For dermoscopy, 50 evaluations (68%) were considered at low risk of bias (Annessi 2007; Arevalo 2008; Argenziano 1998; Argenziano 2011; Benelli 2000; Benelli 2001; Binder 1994; Binder 1995; Binder 1999; Blum 2004; Blum 2004b; Bourne 2012; Carli 2002; Carli 2003; Carli 2003b; Carrera 2016; Dal Pozzo 1999; di Meo 2016; Dolianitis 2005; Dummer 1993; Feci 2015; Ferrari 2015; Friedman 2008; Gereli 2010; Guitera 2009b (Sydney); Kreusch 1992; Lorentzen 1999; Lorentzen 2000; Lorentzen 2008; Malvehy 2014; Menzies 1996; Menzies 2008; Menzies 2013; Nilles 1994; Pagnanelli 2003; Piccolo 2014; Pizzichetta 2002; Pizzichetta 2004; Pupelli 2013; Rao 1997; Rosendahl 2011; Rubegni 2012; Rubegni 2016; Seidenari 2005; Skvara 2005; Stanganelli 2015; Stolz 1994; Tan 2009; Unlu 2014; Zalaudek 2006) and two evaluations were judged high risk, both appearing to report new algorithms or lesion scoring based on their own study data (Blum 2003; Blum 2003b). Twenty-two evaluations (30%) did not provide sufficient information to allow the risk of bias to be fully judged. All studies were judged to have made the diagnosis blinded to the reference standard result; 50 (68%) also clearly reported pre-specification of the diagnostic threshold (40 using named algorithms or pattern analysis, four reporting new algorithms developed using training and test sets (Dal Pozzo 1999; Menzies 1996; Nilles 1994; Stolz 1994) and six providing an indication as to how the diagnosis was to be reached (Annessi 2007; Binder 1995; Carli 2003b; Carrera 2016; Friedman 2008; Lorentzen 1999; Malvehy 2014)).

All 15 image-based visual inspection evaluations were also considered to have made the diagnosis blinded to the reference standard result. hree were considered at low risk of bias due to use of named algorithms with pre-specified thresholds (Benelli 2000; Benelli 2001; Rao 1997) and two (Pizzichetta 2004; Rosendahl 2011) provided some a prior indication as to how the diagnosis was to be reached in the study. The remaining 10 were judged unclear as to pre-specification of the diagnostic thresholds used.

High concern for the applicability of the index tests was recorded for 67 (91%) image-based evaluations of dermoscopy (Alarcon 2014; Annessi 2007; Arevalo 2008; Argenziano 1998; Argenziano 2011; Benelli 2000; Benelli 2001; Binder 1994; Binder 1995; Binder 1999; Blum 2003; Blum 2003b; Blum 2004; Bourne 2012; Carli 2002; Carli 2002b; Carli 2003; Carli 2003b; Carrera 2016; Dal Pozzo 1999; di Meo 2016; Dolianitis 2005; Feci 2015; Ferrari 2015; Ferris 2015; Friedman 2008; Gereli 2010; Gilmore 2010; Glud 2009; Guitera 2009b (Sydney); Hauschild 2014; Kittler 1998; Kittler 2001; Kreusch 1992; Lorentzen 1999; Lorentzen 2000; Lorentzen 2008; Malvehy 2014; Menzies 1996; Menzies 2005; Menzies 2008; Menzies 2013; Nilles 1994; Pagnanelli 2003; Piccolo 2002; Piccolo 2014; Pizzichetta 2002; Rigel 2012; Rubegni 2012; Rubegni 2016; Sboner 2004; Seidenari 1998; Seidenari 2005; Seidenari 2007; Skvara 2005; Stanganelli 1998; Stanganelli 1999; Stanganelli 2005; Stanganelli 2015; Stolz 1994; Tan 2009; Troyanova 2003; Unlu 2014; Wells 2012; Westerhoff 2000; Winkelmann 2016; Zalaudek 2006) (Figure 7), primarily due to blinded interpretation of dermoscopic images without reference to a macro photograph or other patient information (n = 51) or the presentation of average or consensus diagnoses as opposed to for a single observer (n = 35). Twenty five evaluations did not provide sufficient detail regarding the diagnostic threshold used and four were judged to have reported data for non expert observers. The seven evaluations judged as having unclear concern for the applicability of dermoscopy reported data for single observers with the clinical image of the lesion provided alongside the dermoscopic image (Blum 2004b; Lorentzen 2000; Pizzichetta 2004; Pupelli 2013; Rao 1997; Rosendahl 2011; Tenenhaus 2010). All except Tenenhaus 2010 also detailed the diagnostic thresholds used and four clearly described image interpretation by an expert observer (Blum 2004b; Lorentzen 2000; Rosendahl 2011; Tenenhaus 2010).

High concern for the applicability of the index tests was recorded for all 15 visual inspection evaluations due to the image-based nature of test interpretation; only three of these clearly reported diagnosis by a single observer (Pizzichetta 2004; Rao 1997; Rosendahl 2011), the remaining 12 reported average (n = 10) or consensus (n = 2) diagnoses. Thirteen evaluations also did not detail the threshold for diagnosis (all apart from Benelli 2000 and Benelli 2001). Eight evaluations clearly described diagnosis by expert observers (Benelli 2001; Carli 2002b; Carli 2003b; Lorentzen 1999; Rao 1997; Rosendahl 2011; Stanganelli 2005; Troyanova 2003).

Of the 74 included image-based evaluations, 63 (85%) were judged at low risk of bias for the reference standard due to the use of an acceptable reference standard (Alarcon 2014; Annessi 2007; Arevalo 2008; Argenziano 1998; Benelli 2000; Benelli 2001; Binder 1994; Binder 1995; Binder 1999; Blum 2003; Blum 2003b; Blum 2004; Carli 2002; Carli 2002b; Carli 2003; Carli 2003b; Dal Pozzo 1999; di Meo 2016; Dummer 1993; Feci 2015; Ferrari 2015; Ferris 2015; Friedman 2008; Gereli 2010; Gilmore 2010; Glud 2009; Guitera 2009b (Sydney); Hauschild 2014; Kittler 1998; Lorentzen 1999; Lorentzen 2000; Lorentzen 2008; Malvehy 2014; Menzies 1996; Nilles 1994; Pagnanelli 2003; Piccolo 2002; Piccolo 2014; Pizzichetta 2002; Pizzichetta 2004; Pupelli 2013; Rao 1997; Rigel 2012; Rosendahl 2011; Rubegni 2012; Rubegni 2016; Sboner 2004; Seidenari 1998; Seidenari 2005; Seidenari 2007; Skvara 2005; Stanganelli 1998; Stanganelli 1999; Stanganelli 2005; Stanganelli 2015; Stolz 1994; Tan 2009; Troyanova 2003; Unlu 2014; Wells 2012; Westerhoff 2000; Winkelmann 2016; Zalaudek 2006) (Figure 7). Seven evaluations were at high risk of bias, having more than 20% of the benign lesions undergoing follow-up rather than excision (Argenziano 2011; Blum 2004b; Kittler 2001; Menzies 2005) or including some lesions with expert diagnosis only and no follow-up (Bourne 2012; Dolianitis 2005; Menzies 2005; Tenenhaus 2010). Blinding of the reference standard to the index test (in this case the pathology referral diagnosis) was recorded but did not contribute to the overall risk of bias for this domain. Blinding of the reference standard to the original clinical diagnosis was implemented in only one study (Friedman 2008) and was not reported for remainder.

The applicability of the reference standard was: of low concern in 20 evaluations (27%) (all of which reported histopathology interpretation by an experienced histopathologist or by a dermatopathologist) (Alarcon 2014; Annessi 2007; Carli 2003b; di Meo 2016; Ferrari 2015; Ferris 2015; Friedman 2008; Gilmore 2010; Glud 2009; Hauschild 2014; Lorentzen 2000; Malvehy 2014; Piccolo 2002; Pupelli 2013; Rao 1997; Rubegni 2012; Rubegni 2016; Stanganelli 2015; Unlu 2014; Wells 2012); was of high concern in four (5%) (due to the use of expert opinion for classifying the final diagnosis of some lesions) (Bourne 2012; Dolianitis 2005; Menzies 2005; Tenenhaus 2010) and unclear for 50 (68%). In terms of flow and timing, 26 cohorts were judged at high risk of bias (35%) (Alarcon 2014; Arevalo 2008; Argenziano 2011; Blum 2004b; Bourne 2012; Carrera 2016; di Meo 2016; Dolianitis 2005; Dummer 1993; Feci 2015; Ferrari 2015; Guitera 2009b (Sydney); Kittler 2001; Kreusch 1992; Lorentzen 1999; Lorentzen 2008; Malvehy 2014; Menzies 2005; Menzies 2008; Menzies 2013; Pizzichetta 2004; Rosendahl 2011; Stanganelli 2005; Tenenhaus 2010; Westerhoff 2000; Zalaudek 2006), 16 at low risk (22%) (Annessi 2007; Binder 1995; Binder 1999; Blum 2003; Blum 2003b; Blum 2004; Carli 2002b; Ferris 2015; Glud 2009; Hauschild 2014; Lorentzen 2000; Pizzichetta 2002; Rao 1997; Skvara 2005; Stanganelli 2015; Wells 2012) and 32 (43%) did not provide enough information on which to judge this domain (Figure 7). Of those at high risk, 15 evaluations did not use the same reference standard for all participants (differential verification), and 16 did not include all participants in the analysis. Eighteen cohorts (24%) were unclear on the interval between the application of the index test and lesion excision with only 8 (11%) considered to report consecutive diagnosis and excision or biopsy.

Findings

Unless otherwise stated, all analyses were undertaken using HSROC models.

1. Target condition: invasive melanoma and melanocytic intraepidermal variants

Eighty-three study publications reported accuracy data for dermoscopy for the detection of primary target condition - invasive melanoma and intraepidermal melanocytic variants. Two study publications each reported data for two different sets of lesions (Guitera 2009a (Modena); Guitera 2009b (Sydney); Haenssle 2010a (FV); Haenssle 2010b (FU)); and one study (Carli 2002) provided one dataset for in-person dermoscopy and one for image-based interpretation of dermoscopic images. A total of 86 datasets were selected for the primary analyses; 26 for evaluations conducted in-person and 60 for image-based evaluations. Twenty-four of the 83 study publications provided direct comparisons of dermoscopy with visual inspection alone (i.e. data for both tests reported for the same study population). Eleven studies compared in-person visual inspection with in-person visual inspection plus dermoscopy; 11 studies compared diagnosis based on clinical images with diagnosis based on dermoscopic images of the same lesions; and 2 studies compared in-person visual inspection with image-based dermoscopy.

Analyses by clinical pathway and in-person vs image-based design

Summary details of the in-person and image-based studies are provided in Appendix 8 and Appendix 9. Results for the primary analyses are presented in Table 1 and Table 2 with heterogeneity investigations presented in Table 3. Forest plots of study data for each analysis Table 1 are given in Figure 9 and Figure 10; summary estimates for in-person comparisons are depicted in Figure 11 and Figure 12 and for image-based comparisons in Figure 13 and Figure 14.

Clear differences in accuracy were noted between studies undertaken in-person and those which evaluated images, with the accuracy of diagnosis using dermoscopic images and visual inspection of photographs being significantly lower in image-based studies. For dermoscopy, the diagnostic odds ratio for in-person diagnosis was more than 4 times that of image-based diagnosis (RDOR 4.6; 95% CI: 2.4, 9.0, P < 0.001) (Table 3; Figure 15). The high magnitude and importance of this observed difference drive our decision to undertake all analyses separately for in-person and image-based analyses as a primary objective of the review.

Of the 26 evaluations conducted on an in-person basis, 11 contained enough information to describe where on the clinical pathway participants were assessed (coded as 'clear' on pathway), and 15 were considered not to have provided sufficient information to allow the pathway to be identified (coded 'unclear' on pathway). Pathway positions were coded between 1 (test-naïve participants) and 7 (participants identified as high risk for developing melanoma with lesions undergoing follow-up surveillance) (see Figure 3 for diagram of the clinical pathway). For the 60 image-based evaluations, 11 were coded as 'clear' on the pathway and 49 were coded 'unclear'. Across both sets of studies, only 5% (4/86) were considered to have included participants who were presenting for a first structured clinical assessment of a suspicious lesion, the remaining datasets came from studies in participants referred for specialist assessment.

Although there were significant differences between studies undertaken at different points on the pathway, for both in-person (Table 1a; Figure 9) and image-based (Table 1b; Figure 10) approaches there was no clear trend in the estimates of accuracy of dermoscopy according to the degree of prior testing of study participants (as represented by study position on the pathway). Accuracy did appear to be lowest (in terms of DORs) in studies in limited prior testing of the participants (Bourne 2012; Grimaldi 2009; Menzies 2009; Rosendahl 2011) and in those with lesions undergoing follow-up (Haenssle 2010b (FU); Kittler 2001; Skvara 2005; Stanganelli 2015), however, the data were too scarce to draw any firm conclusions. Classification of evaluations by position on the clinical pathway was not further considered analytically.

Dermoscopy added to visual inspection of a skin lesion (in-person evaluations)

Of the 26 in-person evaluations of dermoscopy (Appendix 8 and Figure 9), 11 compared visual inspection alone to visual inspection plus dermoscopy (including two which compared both tests to a CAD-based test (Bono 2002) and one which reported data for a teledermatology consultation (Grimaldi 2009)) and 15 presented data only for dermoscopy in addition to visual inspection (with no data for visual inspection alone), including four which compared in-person dermoscopy to the accuracy of other tests including RCM (Langley 2007; Guitera 2009a (Modena)), exfoliative cytology (Durdu 2011) and CAD (Bauer 2000). Two studies compared the accuracy of different dermoscopy algorithms (Kittler 1999; Menzies 2009).

Two evaluations were conducted in limited prior testing populations (Grimaldi 2009; Menzies 2009). Of those in referred populations, two were considered to have been conducted in participants with equivocal lesions (Carli 2004; Soyer 1995) and one in participants at high risk for developing melanoma with lesions undergoing surveillance (Haenssle 2010b (FU)). The latter study also reported data separately for the same participants at their first visit for lesion assessment (Haenssle 2010a (FV)). Seventeen evaluations were prospective case series, 5 were retrospective (Ahnlide 2016; Bono 2006; Carli 2002; Duff 2001; Stanganelli 2000), and four did not clearly report the design (Bauer 2000; Carli 1994; Gokdemir 2011; Soyer 1995). One study included all melanomas observed across the recruitment period but only a random sample of 50% of observed benign naevi (Guitera 2009a (Modena)). Eighteen evaluations included only pigmented lesions and eight restricted inclusion to lesions considered to be melanocytic in nature. Eighteen of the 26 evaluations (69%) clearly reported including *in situ* melanomas as disease positive, the remaining 8 describing only 'melanomas' not broken down by invasive or *in situ* (Broganelli 2005; Cristofolini 1994; Durdu 2011; Gokdemir 2011; Grimaldi 2009; Morales Callaghan 2008; Nachbar 1994; Stanganelli 2000). The prevalence of invasive melanoma plus atypical intraepidermal melanocytic variants ranged from less than 1% (Haenssle 2010a (FV)) to 41% (Guitera 2009a (Modena); Soyer 1995); median 12% (IQR 5, 21%).

Diagnosis was clearly reported to be conducted on an in-person basis in 24 evaluations (89%) and was assumed in three studies which did not clearly report the use of images or face-to-face diagnosis (Broganelli 2005; Gokdemir 2011; Stanganelli 2000). Diagnosis was recorded by primary care physicians in two studies (7%) (Grimaldi 2009; Menzies 2009), by dermatology residents (trainees) under the supervision of a senior dermatologist (Haenssle 2010b (FU); Haenssle 2010a (FV)) or by a mixed group of dermatology residents and consultants (Ahnlide 2016) in 3 (11%), by dermatologists or presumed to be dermatologists (based on author's institutions) in 17 (63%), by plastic surgeons (Duff 2001) or oncologists (Bono 2002; Bono 2002b; Bono 2006) in four (15%), or was not reported (4%) (Feldmann 1998). Where reported (n=22), the number of observers ranged from 1 to 63 (median 2, IQR 1.25, 4). Test accuracy was reported for a single observer in 48% of evaluations (n=13), for a consensus of two or three observers in 30% (n=8), and this information could not be derived for the remaining 6 evaluations. No formal algorithm to assist diagnosis was reported in 30% of studies (n=8) and 33% (n=9) reported using pattern analysis. The remaining studies used formal algorithms to assist diagnosis: the ABCD algorithm (n=5), the seven-point checklist (n=3), the Menzies criteria (n=1) and seven features for melanoma (n=1) (see Appendix 2 for details of the algorithms used).

Across the 27 evaluations the sensitivity of dermoscopy ranged from 53% to 100% and specificity from 28% to 100% (Figure 9). The low specificities of 28% (Guitera 2009a (Modena) and 56% (Carli 1994) appeared as outliers, all other studies having specificities of 69% or above. Guitera 2009a (Modena) included a relatively high proportion of Spitz naevi in the disease negative group than might be expected in routine clinical practice (19%) while Carli 1994 primarily aimed to distinguish atypical from typical melanocytic lesions and reported accuracy for the decision to excise a lesion as opposed to accurate diagnosis of melanoma.

Results were pooled across algorithms and thresholds as a summary ROC curve (23,487 lesions and 1737 melanomas; Figure 11). Estimates of accuracy obtained from the curve suggest that the specificity of dermoscopy would be 95% (95% CI: 90, 98) at a fixed threshold of 80% sensitivity, and sensitivity would be 92% (95% CI: 87, 95) at a fixed threshold of 80% specificity (Table 2). These 80% fixed values were chosen as they lie within the estimates for the majority of analyses and should only be considered as illustrative examples of the values that might be achieved based on the observed data (see Statistical analysis and data synthesis).

Incremental accuracy from dermoscopy added to in-person visual inspection alone

Of the 13 available in-person evaluations of visual inspection, 11 were reported in these in-person dermoscopy studies and two (Dummer 1993; Unlu 2014) compared in-person visual inspection to image-based dermoscopy (see results for image-based dermoscopy below). Of the 13 evaluations, 77% (n = 10) reported using no algorithm to assist visual inspection diagnosis and three used the ABCD (Stanganelli 2000) or ABCDE (Benelli 1999; Cristofolini 1994) algorithm.

Sensitivities for visual inspection ranged from 38% to 100%; specificities ranged from 45% to 99% (Appendix 10 and Figure 11). The accuracy of visual inspection was compared with the accuracy of dermoscopy estimated from (a) all 26 dermoscopy studies (23,169 lesions and 1664 melanomas) and all 13 in-person visual inspection studies (6740 lesions and 459 melanomas) (Figure 11) and (b) estimated from direct comparisons in the subset of 11 studies that evaluated both visual inspection and dermoscopy on an in-person basis (5854 lesions and 412 melanomas; Figure 12). In both comparisons the accuracy of dermoscopy in addition to visual inspection exceeded that of visual inspection alone (Table 2). In (a) the diagnostic odds ratio (DOR) for dermoscopy was 4.7 (95% CI: 3.0 to 7.5; P < 0.001) times that of visual inspection alone, in (b) it was 4.8 (95% CI: 2.8 to 8.1; P < 0.001) times that of visual inspection alone. These effects correspond to predicted differences in specificity of (a) 20% (95% CI 7, 33%) (based on specificity with dermoscopy of 95% vs 75% for visual inspection) and (b) 21% (95% CI: 2, 39%) (based on specificity with dermoscopy of 96% vs 75% for visual inspection) at a fixed sensitivity of 80% (Table 2) and predicted differences in sensitivity of (a) 16% (95% CI: 8, 23%) (based on sensitivity with dermoscopy of 92% vs 76% for visual inspection) and (b) 15% (95% CI: 7, 23%) (based on sensitivity with dermoscopy of 92% vs 77% for visual inspection) at a fixed specificity of 80% (Table 2).

Dermoscopic images (image-based evaluations)

Of the 60 image-based evaluations of dermoscopy (Appendix 9 and Figure 10), 30 presented data only for dermoscopy, 14 compared diagnosis based on clinical images to diagnosis based on dermoscopic images, 19 compared dermoscopy to the accuracy of other tests including RCM (Alarcon 2014; Ferrari 2015; Guitera 2009b (Sydney); Pupelli 2013; Stanganelli 2015) and CAD based tests (Binder 1994; Blum 2004b; Ferris 2015; Friedman 2008; Glud 2009; Hauschild 2014; Malvehy 2014; Menzies 2005; Piccolo 2002; Piccolo 2014; Rigel 2012; Stanganelli 2005; Wells 2012; Winkelmann 2016). Studies that evaluated dermoscopy images rather than using real-time in-person dermoscopy tended to have been undertaken for reasons of efficiency and not as evaluations of a remote-imaging service, for example, 18 (30%) evaluations compared the accuracy of different dermoscopy algorithms and 13 (22%) compared the accuracy of different observers (see Analyses by observer experience).

Two evaluations recruited participants from limited prior testing populations (Bourne 2012; Rosendahl 2011). Of those in referred populations, nine were considered to have been conducted in participants with equivocal lesions (Alarcon 2014; Annessi 2007; Carli 2003b; Dummer 1993; Ferrari 2015; Kittler 1998; Pupelli 2013; Rubegni 2012; Stolz 1994) and three in participants with lesions undergoing follow-up (Kittler 2001; Skvara 2005; Stanganelli 2015). Seven (12%) evaluations were prospective case series, 33 (55%) were retrospective case series, 17 (28%) used a case-control type design and in (3%) two the design was not clearly reported. All studies prospectively re-interpreted previously acquired dermoscopic images for the purposes of the study. The majority of studies recruited either pigmented (26; 43%) or melanocytic (30; 50%) lesions, including one restricted to melanocytic acral lesions only (Rubegni 2012). Two studies (3%) recruited any lesion selected for excision (Malvehy 2014; Zalaudek 2006) and two (3%) included only amelanotic (Pizzichetta 2004) or amelanotic or hypomelanotic (Menzies 2008) lesions. Forty-four of the 60 evaluations (73%) clearly reported including *in situ* melanomas as disease positive, the remaining 16 describing only 'melanomas' not broken down by invasive or *in situ* (Binder 1994; Binder 1995; Ferrari 2015; Gilmore 2010; Kittler 1998; Malvehy 2014; Pagnanelli 2003; Piccolo 2002; Pizzichetta 2002; Rigel 2012; Rubegni 2016; Seidenari 1998; Seidenari 2005; Stanganelli 1998; Stanganelli 2005; Unlu 2014). The prevalence of disease ranged from 3% (Dummer 1993) to 61% (Stolz 1994) (median 24%, IQR 18, 39%). Prevalence was generally higher in case control type studies (median 37%, IQR 25, 50%) compared to other designs (median 23%, IQR 18% to 33%).

Diagnosis was recorded by dermatologists or assessors presumed to be dermatologists in 80% of studies (n=48), by dermatology residents in one (Carli 2003), and by observers with mixed qualifications in 17% (n=10) including one where all observers were primary care based (three GPS and a clinical nurse in Bourne 2012. Observer qualifications were not reported in one study (Stolz 1994). Where reported (n=56), the number of observers ranged from 1 to 179 (median 3, IQR 2, 8). Test accuracy was reported for a single observer in 42% of evaluations (n=25), for a consensus of two or three observers in 15% (n=9), for a consensus of at least 50% of all observers in one study (Carrera 2016) and for the median or average across observers in 32% of evaluations (n=19); this information could not be derived for the remaining 6 evaluations. Dermoscopic image interpretation was blinded in half of all

evaluations (n=30); the associated clinical (n=17), RCM (n=2) or baseline dermoscopy image (n=1) was provided to assist diagnosis in a further third (n=20). Four evaluations provided information on lesion site, or patient age or gender and the remaining 6 did not describe the provision of additional information. No formal algorithm to assist diagnosis was reported in 38% of studies (n=23) and 32% (n=19) reported using pattern analysis. The remaining 18 studies used formal algorithms to assist diagnosis: the ABCD algorithm (n=6), the seven-point checklist (n=3) or a revised version thereof (n=1), the three point checklist (n=3) the Menzies criteria (n=1) and seven features for melanoma (n=3), or the observers own choice of algorithm (n=1) (Appendix 9).

Across the 60 image-based dermoscopy evaluations, the sensitivity ranged from 22% to 100% and specificity from 31% to 99%. Results were pooled across algorithms and thresholds as a summary ROC curve (13,475 lesions and 2851 melanomas; Figure 13). Estimates of accuracy obtained from the curve suggest that the specificity of dermoscopy would be 82% (95% CI: 75, 87) at a fixed threshold of 80% sensitivity and sensitivity would be 81% (95% CI: 76, 86) at a fixed threshold of 80% specificity (Table 2).

Incremental accuracy of dermoscopic image-based diagnosis compared to visual inspection of images

Of the 11 visual inspection evaluations based on interpretation of clinical images, 82% (n=9) reported using no algorithm to assist image interpretation and two used the ABCD algorithm (Benelli 2000; Benelli 2001). Seven studies reported blinded interpretation of the clinical image with no further patient or lesion information provided, one study allowed observers to view both the clinical and dermoscopic image simultaneously (Pizzichetta 2004), and three did not clearly describe blinding between the clinical and dermoscopic images (Benelli 2000; Stanganelli 2005; Winkelmann 2016).

Sensitivities for image-based visual inspection ranged from 21% to 80%; specificities ranged from 53% to 97% (Figure 13). The accuracy of visual inspection was compared with the accuracy of dermoscopy estimated from (a) all 60 dermoscopy studies (13,475 lesions and 2851 melanomas) and the 11 image-based visual inspection studies (1740 lesions and 305 melanomas) (Figure 13) and estimated from direct comparisons in (b) the subset of 11 studies that evaluated both clinical and dermoscopic images (1740 lesions and 305 melanomas; Figure 14). In both comparisons, the accuracy of dermoscopy exceeded that of visual inspection alone (Table 2). In (a) the diagnostic odds ratio (DOR) for dermoscopy was 5.6 (95% CI: 3.7 to 8.5; P < 0.0001) times that of visual inspection alone, in (b) it was 5.3 (95% CI: 3.5 to 8.0; P < 0.0001) times that of visual inspection alone. These effects correspond to predicted differences in specificity of (a) 40% (95% CI: 27, 57%) (based on specificity with dermoscopy of 82% vs 42% for visual inspection) and (b) 34% (95% CI: 15, 53%) (based on specificity with dermoscopy of 83% vs 48% for visual inspection) at a fixed sensitivity of 80% (Table 2) and predicted differences in sensitivity of (a) 35% (95% CI: 24, 46%) (based on sensitivity with dermoscopy of 81% vs 47% for visual inspection) and (b) 36% (95% CI: 20, 52%) (based on sensitivity with dermoscopy of 83% vs 47% for visual inspection) at a fixed specificity of 80% (Table 2).

Secondary analyses for the detection of invasive melanoma and melanocytic intra-epidermal variants

Covariate investigations

Table 3 and Table 4 report the results of the heterogeneity investigations. Given the large difference in accuracy for in-person evaluations compared to those based on the assessment of dermoscopic images, we elected to undertake all subsequent covariate investigations for in-person (Table 3) and image-based (Table 4) studies separately. In four of the covariate investigations (apart from that by disease prevalence), subgroups with small number of studies were dropped to allow a comparison between the two larger subgroups.

In-person evaluations

Further analysis of the 26 in-person evaluations found no clearly significant relationships between accuracy and the five covariates considered. Some evidence of differences were noted for choice of reference standard and disease prevalence (Table 3).

Choice of reference standard: Observed accuracy was lower in studies that relied on a histological reference standard (n=18), as opposed to those (n=7) that included follow-up of some benign lesions although the difference was not statistically significant (RDOR 0.27; 95% CI: 0.06, 1.22; P = 0.23). Theoretically, the inclusion of a follow-up reference standard has the potential to lower sensitivity (as any melanomas missed on the index clinic visit that are identified on follow-up would be considered as false negatives) and increase specificity (as lesions considered benign and not recommended for excision on the index clinic visit and that do not show any changes on follow-up will increase the number of true negative results). The data observed did demonstrate the anticipated effect on specificity (with specificities at 80% sensitivity increasing from 94% in histology only studies to 99% in histology or follow-up evaluations), however the effect on sensitivity at 80% fixed specificities was the opposite to that anticipated (sensitivity was 6% higher in the histology or follow-up group compared to histology alone). Three of the six in-person evaluations using histology or follow-up as a reference standard reported sensitivities of over 95% (Duff 2001; Dreiseitl 2009; Grimaldi 2009). Of the 9 false negative cases in Duff 2001, 8 melanomas were identified during follow-up (between 5 to 41 months after the initial diagnosis), however with high overall prevalence of disease, sensitivity remained high at 98%. Dreiseitl 2009 and Grimaldi 2009 did not report any melanomas picked up during follow-up (at 1 year and 6 months follow-up, respectively). The perfect sensitivity in Grimaldi 2009 is likely due to lesions classified as positive if they were 'suspicious for melanoma' as opposed to being a likely or definite melanoma. For Dreiseitl 2009 the high sensitivity is likely explained by diagnosis by an expert clinician, and with more than 6 lesions examined per patient assisting diagnosis.

Disease prevalence: Observed accuracy was somewhat higher where disease prevalence of melanoma was 5% or less (RDOR 5.4; 95% CI 0.80, 36.6), and where prevalence was greater than 20% (RDOR 5.0; 95% CI 0.78, 32.4) compared to those with disease prevalence between 5% and 10% (LR test for differences between groups: P = 0.008). No obvious explanation for these results could be derived from the study characteristics (Table 3).

Other investigations: The RDOR for use of no algorithm to aid diagnosis compared to a named algorithm was 1.4 (95% CI 0.34, 5.6; P = 0.17), for a single observer compared to a consensus of two or more observers was 1.0 (95% CI 0.18, 5.8; P = 0.30), and for evaluations including only melanocytic lesions compared to any pigmented lesion was 0.48 (95% CI 0.12, 2.0; P = 0.60) (Table 3).

Image-based evaluations

For the 60 image-based evaluations, no clearly significant relationships were noted between accuracy and the five covariates. The choice of reference standard showed an effect in the opposite direction to that observed for the in-person evaluation (Table 4). Observed accuracy was higher in studies that relied on a histological reference standard (n=48), as opposed to those (n=8) that included follow-up of some benign lesions (RDOR 2.8; 95% CI 0.92, 8.9; P = 0.19). At a fixed specificity of 80%, observed sensitivity in studies using a follow-up reference standard was lower (65%) compared to those using histology alone (84%) as might be expected, however at a fixed sensitivity of 80%, specificities in studies that included follow-up of some benign lesions was also lower (64%) compared to those using histology alone (84%). This effect is likely due to a combination of reasons that cannot be derived from the data due to heterogeneity in participants, tests and observers.

Disease prevalence: Disease prevalence was higher in image based studies than in person studies and a different grouping for prevalence was used. Observed accuracy appeared highest where disease prevalence of melanoma was 20% or less (RDOR 30.7; 95% CI 1.51, 6.24) compared to prevalence >20-30% and higher.

For the other characteristics investigated, similar results to those obtained for in-person evaluations were observed for use of a named algorithm, the effect from restriction to melanocytic lesions only was in the opposite direction although non-significant (P = 0.16), and results to a greater order of magnitude were observed for diagnosis by a single observer compared to a consensus of two or more observers.

1.1.1. Analyses by algorithms used to assist dermoscopy

Details of the algorithms used to assist diagnosis are provided in Appendix 2. Results by algorithm used (or not used) are reported in Table 5 for each of the target conditions under consideration in this review. All analyses in this section were undertaken using the bivariate normal model.

In-person evaluations of dermoscopy added to visual inspection

The 26 in-person evaluations of dermoscopy added to visual inspection of a lesion provide a total of 40 datasets using different algorithms or diagnostic thresholds for the detection of invasive melanoma or atypical intraepidermal melanocytic variants. Nine of the datasets did not report the use of any algorithm to assist diagnosis, 8 reported data for pattern analysis, and the remaining 23 datasets used one or more of 7 different formally developed algorithms (Table 5a).

A pooled sensitivity of 88% (95% CI 75, 95%) and specificity of 87% (95% CI 80, 92%) was estimated for observer diagnosis without the use of a formal algorithm (n=8 datasets; 4707 lesions, 849 melanomas). The approach to diagnosis was not well described; however, most studies in this dataset reported accuracy for the clinician's correct diagnosis of melanoma rather than the decision to biopsy or excise a lesion (Appendix 8). Pooled results for studies using pattern analysis were similar but with narrower confidence intervals for sensitivity (sensitivity 92%, 95% CI 87, 95%; specificity 92%, 95% CI 68, 98%; 6 datasets with 4307 lesions and 296 melanomas).

Of the more formal algorithms for melanoma diagnosis, results could be pooled for only two. Five datasets (1438 lesions and 160 melanomas) using the ABCD algorithm at a threshold of >5.45 produced a sensitivity of 81% (95% CI 62, 92%) and specificity of 92% (95% CI 82, 97). Two evaluations (11,137 lesions and 127 melanomas) reported data for the seven point checklist (7PCL) at a threshold of >=3, giving a sensitivity of 67% (95% CI 46, 83) and specificity of 96% (95% CI 88, 99%). The latter result is based on a single study publication which reports results for 8449 lesions detected on a patient's first clinic visit (Haenssle 2010a (FV)) and separately for 2373 lesions examined during follow-up (Haenssle 2010b (FU)). The ABCDE algorithm, seven features for melanoma (7FFM) and the Menzies criteria were each assessed in a single study on an in-person basis; results were generally similar to those observed above (Table 5a).

Image-based evaluations of dermoscopic images

The 60 evaluations of dermoscopic images provide a total of 113 datasets using different algorithms or diagnostic thresholds for the detection of invasive melanoma or atypical intraepidermal melanocytic variants. Twenty-eight of the datasets did not report the use of any algorithm to assist diagnosis (4 studies reporting data at two thresholds*)*, 22 report data for pattern analysis (2 studies reporting data at two thresholds*)*, and the remaining 63 datasets used one or more of 14 different formally developed algorithms (Table 5a).

For observer diagnosis without the use of a formal algorithm, diagnostic thresholds (i.e. the clinical decision that was recorded by the clinician concerned) were poorly reported; however, we attempted to differentiate between those studies reporting the observer's correct diagnosis of melanoma from those reporting the decision to excise a lesion. Pooling all data regardless of threshold (24 datasets; 4498 lesions and 941 melanomas) gave a sensitivity of 76% (95% CI 70, 82%) and specificity of 79% (95% CI 71, 85%). Restricting the analysis to the 18 datasets reporting data for observers correctly diagnosing melanoma (4118 lesions; 795 melanomas) gave a sensitivity of 77% (95% CI 69, 83) and specificity of 84% (95% CI 76, 89%). For the 10 datasets that reported data for the decision to excise a lesion (831 lesions; 263 melanomas), sensitivity was similar at 79% (95% CI 69, 86%) but specificity reduced to 55% (95% CI 50, 61%). Pooled results for 20 evaluations (4621 lesions and 989 melanomas) reporting use of pattern analysis resulted in higher sensitivity (83%, 95% CI 76, 88%) and specificity (87%, 95% CI 80, 92%) compared to the no algorithm reported studies but results were both lower in comparison to the in-person valuations.

Sufficient data were available to allow pooling for seven different formal algorithms to assist diagnosis (Table 5a); all summary estimates showed either lower sensitivity or lower specificity, or both, in comparison to either the no algorithm or pattern analysis datasets. The ABCD checklist at a threshold of > 5.45 (7 datasets; 2471 lesions and 406 melanomas) had a sensitivity of 81% (95% CI 60, 92%) and specificity of 81% (95% CI 69, 89%). At the lower threshold of > 4.75 for diagnosis of melanoma, sensitivity remained at 81% (95% CI 67, 90%) with narrower confidence intervals with a lower specificity of 72% (95% CI 93, 80%) (10 datasets; 4242 lesions and 816 cases).

The 7PCL was evaluated in 11 datasets at a threshold of >= 3 (3408 lesions and 798 melanomas), pooled sensitivity was 80% (95% CI 63, 91%) and specificity 67% (95% CI 51, 80). Four evaluations that did not report the threshold used with the 7PCL demonstrated lower sensitivity (72%, 95% CI 56, 84%) but higher specificity (79%, 95% CI 61, 90).

The 7FFM tool was assessed in four datasets with 2200 lesions and 340 melanomas, sensitivity was 89% (95%CI 76, 96%) with specificity 84% (95%CI 78, 89%). The Menzies criteria was evaluated in four datasets using the method described in the original Menzies 1996 paper, pooled sensitivity was 78% (95% CI 38, 96) and specificity 63% (95% CI 39, 81) (1856 lesions and 317 melanomas).

Seven evaluations of the 3PCL at a threshold of >= 2 were pooled (1505 lesions and 363 melanomas), summary sensitivity was 74% (95% CI 61, 85) and specificity 60% (95% CI 42, 76%). Sixteen additional datasets reporting data for other algorithms or at different thresholds are reported in Table 5a, however study numbers are too small to describe results in any detail.

1.1.2. Analyses by observer experience and qualifications

Table 6 and Table 7 report results for the effect of observer experience and qualifications. Observer experience was generally poorly described in the study reports (see Appendix 8 and Appendix 9), however we attempted broad classifications by expertise in dermoscopy and reported qualifications with the 'consultant' category in the latter analysis separated into 'Expert consultant' (for any study describing observers as expert or experienced) and 'Consultant' where experience or expertise was not otherwise reported (for example for those that described observers as dermatologists). Results are described separately for in-person (Figure 16; Figure 17) and image-based evaluations (Figure 18; Figure 19).

For the in-person evaluations, the majority of observers were classified as having high dermoscopy experience (n= 14) or as experience not reported (n=10). Two studies reported data for GPs provided with some dermoscopy training for the purposes of the study (Grimaldi 2009; Menzies 2009). No statistically significant differences were found between groups (Table 6) although the poorest performance was noted in the GP training group.

The 60 image-based evaluations provided 77 datasets according to observer experience; 13 evaluations providing data for more than one observer (Argenziano 1998; Benelli 2001; Binder 1995; Ferris 2015; Hauschild 2014; Menzies 2005; Pagnanelli 2003; Piccolo 2002; Piccolo 2014; Seidenari 1998; Seidenari 2005; Stanganelli 1999; Tan 2009). The LR test for differences between groups was statistically significant (P < 0.001). Using the high experience group as the reference (34 datasets; 8933 lesions and 1956 melanomas), the RDOR for the observers where experience was not reported (11 datasets; 2777 lesions and 465 melanomas) was 2.0 (95% CI 0.8, 4.9), while the RDORs for the lower experience groups all suggested lower accuracy (Table 6; Figure 17). The RDORs for each group in comparison to the high experience group were: moderate experience 0.64 (95% CI 0.37, 1.1; 5 datasets, 678 lesions and 193 melanomas); 'low' experience 0.30 (95% CI 0.15, 0.58; 6 datasets; 448 lesions and 123 melanomas); for the 'mixed' experience group was 0.25 (95% CI 0.07, 0.81; 5 datasets, 473 lesions and 117 melanomas); and for the 'trained' group was 0.51 (95% CI 0.25, 1.02; 11 datasets, 1087 lesions and 240 melanomas).

Similar trends were observed when evaluations were sub-grouped according to reported observer qualifications, however data for clinicians other than consultant or 'consultant experts' were relatively sparse, especially for the in-person evaluations where no statistically significant differences between groups was determined (Table 7; Figure 18). For the image-based evaluations accuracy was highest for the 'Expert consultant' group (DOR 19.4, 95% CI 13.1, 28.8; 33 datasets, 8664 lesions and 1854 melanomas) (Figure 19). Relative DORs in comparison to the 'expert' group were 0.61 (95% CI 0.40, 0.92) for observers described as 'dermatologists' (25 datasets; 4589 lesions and 955 melanomas), 0.31 (95% CI 0.14, 0.71 for registrar (trainee) or resident level observers (5 datasets; 927 lesions and 138 observers), and 0.10 (95% CI 0.04, 0.25) for the GP group (3 datasets; 288 lesions and 55 melanomas). Results for the GP group may simply be attributed to small sample sizes; however, the lowest sensitivity for detection of melanoma (22%) was observed for Bourne 2012 in which seven of nine included melanomas *in situ* or lentigo maligna and lowest specificity (44%) was observed for Piccolo 2014 which included a relatively high percentage of Spitz naevi (14% of the disease negative group) which may have been more difficult to differentiate from melanomas. Both studies also implemented blinded dermoscopy image interpretation whereas the third study in this group (Menzies 2005) also provided the clinical image and information on patient history to the interpreting clinicians.

1.1.3. Results of sensitivity analyses

A number of sensitivity analyses were planned in our generic protocol. One, restricting comparisons between dermoscopy and visual inspection alone to studies where both tests have been evaluated in the same study (direct comparisons), was discussed alongside the main test comparisons above (Table 2). For completeness, the results of these are included in Table 8 (in-person evaluations) and Table 9 (image-based evaluations) along with the results of all other sensitivity analyses.

In-person evaluations

Analyses restricting studies to those avoiding partial verification (including only those which allowed histology or follow-up) increased the relative benefit from adding dermoscopy to visual inspection from an RDOR of 4.7 (95% CI 3.0, 7.5) to 14.4 (95% CI 4.4, 7.6) however study numbers were small and the increase in sensitivity at 80% specificity and in specificity at 80% sensitivity remained similar (Table 8). Limited differences were observed for the analyses restricting studies to those with low risk of bias for the index test or low risk of bias for the reference standard. An additional *post hoc* analysis restricting studies to those with low risk of bias for flow and timing resulted in small study numbers and did not appear to have a large impact on accuracy. Planned analyses restricting to studies with at least a three month interval between the index test and the reference standard was at least three months, and where concerns around applicability for participant selection are low, were not possible due to lack of studies.

Image-based evaluations

Sensitivity analyses for image-based evaluations were more difficult to interpret in terms of the differences between diagnosis using dermoscopic images versus visual inspection of images due to small study numbers for visual inspection in comparison to the numbers for dermoscopy (Table 9), e.g. for restriction to those which allowed histology or follow-up as a reference standard (7 datasets for dermoscopy compared to 0 for visual inspection), for low risk of bias for the index test (40 for dermoscopy vs 3 for visual inspection) and for low risk of bias for flow and timing (11 datasets for dermoscopy vs 1 for visual inspection). Restriction to studies with low risk of bias for the reference standard made very little difference to the accuracy of either test or to the RDOR for dermoscopy versus visual inspection.

Planned analyses restricting to studies with at least a three month interval between the index test and the reference standard was at least three months and where concerns around applicability for participant selection are low were again not possible due to lack of studies.

An additional *post hoc* sensitivity analysis restricting studies to those which did not use a case-control design increased the accuracy of visual inspection of images from a DOR of 3.2 (95% CI 1.9, 5.4) to DOR 7.2 (95% CI 3.5, 14.8) for the 7 remaining datasets, and increased the DOR for diagnosis using dermoscopic images from 17.8 (95% CI 12.3, 25.7) to 24.3 (95% CI 15.2, 39.0) for the remaining 37 datasets; the RDOR between tests reduced from 5.6 (95% CI 3.7, 78.5) to 3.4 (95% CI 1.8, 6.4) (Table 9). From the sensitivities and specificities estimated from SROC curves, this fall in RDOR appears to be primarily related to an increase in accuracy for diagnosis based on visual inspection of images rather than a fall in accuracy for dermoscopic due to the exclusion of case-control studies. The direction of this finding is contrary to the standard expectation that case-control studies over-estimate test accuracy compared to other designs (Rutjes 2006).

2. Target condition: invasive melanoma only

In this section we present the results for studies of dermoscopy for the identification of invasive melanoma, according to the approach taken for diagnosis: in-person or image-based evaluations. Summary characteristics of studies are presented in Appendix 11, with forest plots of study data in Appendix 12 and results of meta-analyses in Table 10 and Figure 20 and Figure 21.

Dermoscopy added to visual inspection of a skin lesion (in-person evaluations)

Six studies evaluated the accuracy of in-person dermoscopy for the detection of invasive melanoma only, one of which also reported data for the primary target condition (Feldmann 1998) and two of which presented data for visual inspection (Krahn 1998; Viglizzo 2004). All studies were case series based in secondary care or specialist units apart from Coras 2003 which was based in a private dermatology clinic. All recruited participants with pigmented lesions, Viglizzo 2004 restricting to melanocytic lesions only. Four studies did not report using any formal algorithm to assist dermoscopy diagnosis (Coras 2003; Krahn 1998; Piccolo 2000; Viglizzo 2004); Feldmann 1998 used the ABCD checklist and Ascierto 2010 a modified version of the Kenet risk stratification approach (referenced to Ascierto 1998). The prevalence of melanoma ranged from 5% (Feldmann 1998) to 49% (Krahn 1998). All studies used a histological reference standard.

The sensitivity of in-person dermoscopy ranged from 64% to 100% and specificities ranged from 93% to 98% (Appendix 12). In meta-analysis the DOR was 129 (95% CI 19.2, 870) (789 lesions and 115 melanomas). The specificity of in-person dermoscopy at 80% fixed sensitivity was 97% (95% CI: 94, 98) and sensitivity at 80% fixed specificity was also 97% (95% CI 46, 100) (Table 10). Again, these sensitivities and specificities at fixed values should be taken as illustrative of the data observed.

In Feldmann 1998, the sensitivity for the detection of invasive melanoma alone was 11% higher compared to sensitivity for the detection of invasive melanoma or atypical intraepidermal melanocytic variants (64% vs 53%) because the 5 included melanoma *in situ* lesions were all classified as negative for melanoma on dermoscopy and were classed as true negative results for the detection of invasive melanoma alone.

Incremental accuracy from dermoscopy added to in-person visual inspection alone

The two studies providing direct comparisons of visual inspection alone and visual inspection plus dermoscopy reported using no algorithm to assist visual inspection diagnosis (Krahn 1998; Viglizzo 2004). Observers in both studies were assumed to be dermatologists based on reported authors' institutions.

Sensitivities for visual inspection were 79% (Krahn 1998) and 67% (Viglizzo 2004); specificities were 78% and 95%, respectively (Appendix 12). The accuracy of visual inspection was compared with the accuracy of dermoscopy estimated from (a) all 6 dermoscopy studies (789 lesions and 115 melanomas) and both in-person visual inspection studies (147 lesions and 51 melanomas) and estimated from (b) direct comparisons in the subset of 2 studies that evaluated both visual inspection and dermoscopy on an in-person basis (147 lesions and 51 melanomas). In both comparisons the accuracy of dermoscopy added to visual inspection exceeded that of visual inspection alone (Table 10). In (a) the diagnostic odds ratio (DOR) for dermoscopy was 6.2 (95% CI: 1.5 to 26.6; P = 0.015) times that of visual inspection alone, in (b) it was 11.3 (95% CI: 1.4 to 89.8; P = 0.015) times that of visual inspection alone. These effects correspond to predicted differences in specificity of (a) 13% (95% CI: -1, 27%) (based on sensitivity with dermoscopy of 97% vs 84% for visual inspection) and (b) 24% (95% CI: -21, 69%) (based on sensitivity with dermoscopy of 99% vs 75% for visual inspection) at a fixed sensitivity of 80% (Table 10); and predicted differences in sensitivity of (a) 13% (95% CI: -0, 27%) (based on specificity with dermoscopy of 97% vs 84% for visual inspection) and (b) 15% (95% CI: 2, 29%) (based on specificity with dermoscopy of 94% vs 78% for visual inspection) at a fixed specificity of 80% (Table 10).

Dermoscopic images (image-based evaluations)

Thirteen datasets reported the accuracy of image-based dermoscopy for the detection of invasive melanoma, none of which reported data for the primary target condition. Eight evaluations included series of lesions observed in secondary care or specialist clinic settings (prevalence 10% to 36%). The remaining five evaluations used a case control type design, with separate sampling of melanoma and benign lesion images (prevalence ranged from 27% to 65%). Studies used the ABCD checklist (Lorentzen 2000; Menzies 2013), the Menzies

algorithm (Arevalo 2008; Menzies 1996; Westerhoff 2000) or their own algorithm (Kreusch 1992; Nilles 1994) to assist dermoscopic diagnosis. Six evaluations did not report using any algorithm to assist diagnosis.

Five evaluations presented only the dermoscopic image with no further patient information (Arevalo 2008; Lorentzen 2008; Menzies 1996; Nilles 1994; Troyanaova 1993), five presented observers with a concurrent clinical image of the lesion (Hauschild 2014; Lorentzen 1999; Lorentzen 2000; Rao 1997; Westerhoff 2000); two provided only lesion site (Kreusch 1992) or site, age and gender (Friedman 2008), and one did not describe any further information (Menzies 2013). Images were interpreted by dermatologists or assumed to be by dermatologists in ten studies, by dermatologists or melanoma fellows in Rao 1997, by GPs in Westerhoff 2000 and by mixed secondary care clinicians in Friedman 2008.

Sensitivities ranged from 48% to 100%; specificities ranged from 49% to 97% (Appendix 12). In meta-analysis the DOR was 27.5 (95% CI: 12.2, 61.7) (5618 lesions and 1092 melanoma cases). Specificity at 80% fixed sensitivity was 87% (95% CI 75, 94%) and sensitivity at 80% fixed specificity was 88% (95% CI 75, 94%) (Table 10).

Incremental accuracy of dermoscopic image-based diagnosis compared to visual inspection of images

The four studies providing direct comparisons of diagnosis based on clinical images and diagnosis based on dermoscopic images reported using no algorithm to assist visual inspection diagnosis (Lorentzen 1999; Troyanova 2003; Westerhoff 2000) or use of the ABCD algorithm (Rao 1997). Observers were dermatologists (Lorentzen 1999; Troyanova 2003), a melanoma fellow (Rao 1997) or GPs (Westerhoff 2000).

Sensitivities for visual inspection ranged from 62% to 86%; and specificities from 54% to 89%, respectively (Appendix 12). The accuracy of visual inspection was compared with the accuracy of dermoscopy estimated from (a) all 13 dermoscopy studies (5618 lesions and 1092 melanomas) and the four visual inspection studies (454 lesions and 145 melanomas) and estimated from direct comparisons in (b) with the subset of four studies that evaluated both visual inspection and dermoscopy on an image-based basis (454 lesions and 145 melanomas). In both comparisons the accuracy of diagnosis based on dermoscopic images exceeded that based on clinical images (Table 10). In (a) the diagnostic odds ratio (DOR) for dermoscopy was 2.5 (95% CI: 1.2 to 5.1; P = 0.032) times that of visual inspection alone, in (b) it was 3.4 (95% CI: 1.0 to 11.1; P = 0.049) times that of visual inspection alone. These effects correspond to predicted differences in specificity of (a) 13% (95% CI -1, 28%) (based on sensitivity with dermoscopy of 87% vs 74% for visual inspection) and (b) 44% (95% CI -20, 100%) (based on sensitivity with dermoscopy of 89% vs 45% for visual inspection) at a fixed sensitivity of 80% (Table 10) and predicted differences in sensitivity of (a) 15% (95% CI -1, 30%) (based on specificity with dermoscopy of 88% vs 72% for visual inspection) and (b) 11% (95% CI 1, 22%) (based on specificity with dermoscopy of 83% vs 72% for visual inspection) at a fixed specificity of 80% (Table 10).

3. Target condition: any skin lesion requiring excision

In this section we present the results for studies of visual inspection for the identification of any skin lesion requiring excision (for each study data could only be extracted for the detection of any skin cancer), according to the approach taken for diagnosis: in-person or image-based evaluations. Summary characteristics of studies are presented in Appendix 13, with forest plots of study data in Figure 22 and Figure 23 and results of meta-analyses in Table 11. Heterogeneity was too high and data too sparse to make formal statistical comparisons between tests, thus the analysis focuses on describting the observed accuracy. Only meta-analytical models assuming underlying symmetric SROC curves could be fitted to these data.

Dermoscopy added to visual inspection of a skin lesion (in-person evaluations)

Four datasets evaluated the accuracy of in-person dermoscopy for the detection of any skin lesion requiring excision (Argenziano 2006; Durdu 2011; Stanganelli 2000; Soyer 2004), one of which also reported data for the primary outcome (Durdu 2011) and two reported data for visual inspection alone (Argenziano 2006; Stanganelli 2000). Studies were based in primary care (with diagnosis by GPs) (Argenziano 2006) or secondary care or specialist referral clinics with diagnosis by dermatologists. The prevalence of skin cancer ranged from 3% (Stanganelli 2000) to 51% (Argenziano 2006). Studies used the ABCD algorithm (Durdu 2011), the 3PCL (Argenziano 2006), pattern analysis (Stanganelli 2000) or no algorithm (Soyer 2004) to assist diagnosis. Stanganelli 2000 supplemented a histological reference standard with clinical follow-up, and the others reported data compared to histology alone.

Sensitivities ranged from 85% to 98%; specificities ranged from 26% to 100% (Figure 22). In meta-analysis the DOR was 232 (95% CI: 16.0, 3354) (3880 lesions and 260 skin cancer cases). No formal comparison with in-person visual inspection could be made due to heterogeneity and sparsity of data; however, the DOR for the two studies reporting data for visual inspection alone (3457 lesions and 151 skin cancers) was 15.0 (95% CI: 0.18, 1225) (Argenziano 2006; Stanganelli 2000), compared to 88.1 (95% CI: 1.1, 7338) for in-person dermoscopy in these same two studies (3449 lesions and 137 skin cancers; the total number of lesions and melanomas differs because Argenziano 2006 was a between person comparison study with a different number of lesions randomised to each arm). Sensitivities at 80% fixed specificity and specificities at 80% fixed sensitivity were both 17% higher using dermoscopy (both 96% with dermoscopy compared to 79% for visual inspection alone) due to the use of symmetric ROC curves for these analyses.

The lowest sensitivity and specificity for dermoscopy were observed in Argenziano 2006, however 2x2 data for the GP diagnosis using the 3PCL could only be included for the 77 lesions selected for excision by an expert dermatologist as the remaining 1126 for which GP diagnosis was recorded did not have an adequate reference standard for inclusion in our review. In Durdu 2011 specificity estimates were not affected by the wider definition of the target condition, however sensitivity increased from 80% for detection of melanoma or atypical intraepidermal melanocytic variants to 98% for detection of any lesion requiring excision, as all 34 BCCs were correctly identified.

Dermoscopic images (image-based evaluations)

Five datasets reported the accuracy of image-based visual inspection for the detection of any skin lesion requiring excision (Carli 2002b; Lorentzen 2008; Rosendahl 2011; Stanganelli 1998; Zalaudek 2006), all of which also reported data for the primary target condition or for the detection of invasive melanoma alone (Lorentzen 2008) and three reported data for diagnosis based on clinical images (Carli 2002b; Rosendahl 2011; Stanganelli 1998). Studies selected images from secondary care clinics or specialist units (Carli 2002b; Lorentzen 2008; Stanganelli 1998), one from a primary care practice (Rosendahl 2011) and was not clearly reported in Stanganelli 1998. The prevalence of lesions suitable for excision ranged from 22% (Rosendahl 2011) to 47% (Stanganelli 1998); the latter selecting images for use in a dermoscopy training study. Diagnosis was based on the 3PCL (Zalaudek 2006), pattern analysis (Rosendahl 2011) or no formal algorithm. Data were presented for a single dermatologist (Rosendahl 2011), for a consensus of two dermatologists (Carli 2002b), for the average across 20 dermatologists (Stanganelli 1998) or 150 dermatologists (Zalaudek 2006) or was not clearly reported (Lorentzen 2008). Observers were also provided with the clinical image for the same lesion (Rosendahl 2011; Stanganelli 1998), with lesion site, and patient age and gender (Zalaudek 2006) or with no further clinical information to assist diagnosis (Carli 2002b; Lorentzen 2008).

Sensitivities ranged from 78% to 100%; specificities ranged from 72% to 96% (Figure 23). In meta-analysis the DOR was 37.5 (95% CI 8.8, 161) (815 lesions and 217 skin cancer cases). No formal comparison with diagnosis based on clinical images could be made due to heterogeneity and sparsity of data, however the DOR for the three studies reporting image-based visual inspection (547 lesions and 138 skin cancers) was 12.1 (95% CI 5.4, 26.7) (Carli 2002b; Rosendahl 2011; Stanganelli 1998), compared to 18.4 (95% CI 8.1, 41.7) for image-based dermoscopy in these same 3 studies. Sensitivities at 80% fixed specificity and specificities at 80% fixed sensitivity were 7% higher using dermoscopy (both 82% with dermoscopy compared to 75% for visual inspection of clinical images).

The wider definition of the target condition to include any skin lesion requiring excision led to increased sensitivities and lower specificities in three studies due to classification of BCCs as true positives rather than false negatives (Carli 2002b; Rosendahl 2011; Stanganelli 1998). Data for Rosendahl 2011 and Stanganelli 1998 were also extracted for the correct diagnosis of any malignancy rather than correct diagnosis of each individual type of skin cancer, which led to considerable increased in sensitivity in both studies.

4. Evaluations of dermoscopy training

Observer accuracy using dermoscopy was evaluated before and after a dermoscopy training intervention in six studies; two reported data for detection of invasive melanoma alone (Troyanova 2003; Westerhoff 2000) and four reported data for the detection of invasive melanoma and atypical intraepidermal melanocytic variants (Pagnanelli 2003; Piccolo 2014; Stanganelli 1999; Tan 2009). A further 14 studies reported the delivery of some form of dermoscopy training either prior to the study commencing (Kittler 1998; Seidenari 2007) or within the context of the study itself (Argenziano 1998; Argenziano 2006; Binder 1999; Carli 2003; Dolianitis 2005; Grimaldi 2009; Kittler 1998; Menzies 2008; Menzies 2009; Seidenari 2007; Stanganelli 1998; Zalaudek 2006). Six of the latter group of studies compared the accuracy of diagnosis based on visual inspection alone (pre-dermoscopy training) to visual inspection and dermoscopy (post-dermoscopy training); these data are incorporated into the visual inspection versus dermoscopy comparisons reported above.

Details of the training interventions provided in the six eligible studies are provided in Appendix 14. Results of the analyses are reported in Table 12 and Figure 24. All studies were image-based evaluations.

For the detection of invasive melanoma or atypical intraepidermal melanocytic variants, all four evaluations (Pagnanelli 2003; Piccolo 2014; Stanganelli 1999; Tan 2009) demonstrated an increase in the average sensitivity of dermoscopy of between 13 and 15% (pre-training sensitivity ranged from 36% to 80% and post-training from 73% to 93%). No change in average specificity was observed following dermoscopy training for three studies and specificity fell from 70% pre-training to 44% post-training in Piccolo 2014. The pooled analysis showed no impact accuracy from dermoscopy training (RDOR 1.4, 95% CI 0.38, 5.3).

Three of the four studies reported the training of dermatologists (n = 83 in Stanganelli 2000) or of a mixed group of dermatologists, registrars or residents (n = 16 in Pagnanelli 2003; n = 6 in Tan 2009). Pagnanelli 2003 also included 3 medical students in their group of '16 trainees'. These three studies provided web-based interactive training (Pagnanelli 2003; Tan 2009), with an expectation of a time commitment of 1 hour per day for two weeks (Pagnanelli 2003) or with a dermatoscope provided for use in clinical practice for 10 months between tests (Tan 2009), or in-person dermoscopy training workshops (Stanganelli 1999). In Piccolo 2014 however, the 'trainee' was a single GP who undertook similar online training using an interactive atlas of dermoscopy, which may explain the outlying result for specificity.

For the detection of invasive melanoma, both evaluations demonstrated an increase in the average sensitivity of dermoscopy in the order of 16 to 18% following dermoscopy training (from 76% to 92% in Troyanova 2003's study of 32 dermatologists and from 58% to 76% in Westerhoff 2000's study of 74 GPs), with minimal impact on specificity (84% before and after training in Troyanova 2003 and 56% before and 58% after training in Westerhoff 2000). The pooled analysis showed a non statistically significant increase in accuracy after training of 3.1 times that before training (95% CI 0.94, 10.6, P = 0.06; 150 lesions and 75 cases). As well as the differences in clinician qualifications, the content and duration of the training programmes also varied. Troyanova 2003 provided 6 hours of in-person teaching daily for two consecutive days; the test using clinical and dermoscopic images of 50 lesions was undertaken at the beginning and at the end of the course. In Westerhoff 2000, GPs were provided with a pictorial atlas outlining the Menzies approach to dermoscopic diagnosis and given a one hour presentation on the method; the pre- and post-tests were undertaken at the leisure of the individual GPs.

Discussion

Summary of main results

Dermoscopy to assist the diagnosis of melanoma has been evaluated in a range of study populations, on an in-person basis added to visual inspection of a skin lesion and using dermoscopic images, and both with and without the use of published algorithms to assist diagnosis. Wide variations in both sensitivity and specificity for dermoscopy use were observed for all definitions of the target condition. In terms of methodological quality, many studies were at high or unclear risk of bias for participant selection and for timing of diagnosis in

relation to reference standard diagnosis, but were at low risk of bias for the index test and reference standard. Concern around the applicability of studies was almost universally poor due to restricted inclusion of lesions (for example inclusion of only melanocytic lesions or of lesions selected for excision based on the clinical or dermoscopic diagnosis), and lack of reproducibility of diagnostic thresholds. Poor reporting in the primary studies hindered attempts to analyse studies according to their position on the clinical pathway and to fully assess sources of heterogeneity and methodological quality.

In this review we have estimated the incremental accuracy of dermoscopy in comparison to visual inspection using summary ROC curves rather than by estimating average sensitivity and specificity operating points. We have reported points from the fitted SROC curves (the sensitivity at 80% specificity, and the specificity at 80% sensitivity), however these are for illustrative purposes and should not be quoted as the actual performance of dermoscopy. Whilst it may not be possible to estimate the absolute accuracy of dermoscopy, nor to make any clear recommendations to ensure that dermoscopy is used in such a way as to maximise sensitivity, we can make a strong comparison between dermoscopy and visual inspection alone despite the limitations and heterogeneity of included studies, particularly from the studies which make within patient comparisons between diagnostic strategies of visual inspection alone, and visual inspection supplemented by dermoscopy. We also present results separately for in-person and image-based studies, as we observed clear differences in their findings. We choose to emphasise the in-person findings over the image-based studies as these are more applicable to typical practice.

Thus whilst we cannot answer the overall question of how accurate dermoscopy is, we are able to assess the incremental gain in accuracy of using dermoscopy, and identify some characteristics which increase or decrease its accuracy.

Five main findings can be drawn from our review:

1) On average, the addition of dermoscopy to in-person visual inspection of a lesion increases both sensitivity and specificity by a considerable margin.

Approximately one third of eligible studies presented data for in-person dermoscopy (26/86) for the primary target condition of invasive melanoma or atypical intraepidermal melanocytic variants. A range of study populations were included and a number of different algorithms to assist interpretation were employed, such that considerable heterogeneity in both sensitivity and specificity was observed for both visual inspection alone and for visual inspection plus dermoscopy. The Summary of findings table 1 presents key results and translates summary estimates to a hypothetical cohort of 1000 lesions.

Sensitivity: At a fixed specificity of 80%, the use of dermoscopy increased the sensitivity of in-person visual inspection by 26%, from 76% to 92%. Assuming melanoma or atypical intraepidermal melanocytic variant prevalences of 5%, 12% and 21%, a test sensitivity of 92% with the added use of dermoscopy would reduce the number of melanomas missed in comparison to using visual inspection alone by 8, 19 and 33 (resulting in 4, 10 and 17 melanomas missed). An assumed test specificity of 80% (for both visual inspection and visual inspection plus dermoscopy) would result in 190, 176 and 158 false positive test results (or unnecessary excisions).

Specificity: At a fixed sensitivity of 80%, the use of dermoscopy increased the specificity of in-person visual inspection by 20%, from 75% to 95%. Applying these results to a cohort of 1000 lesions at the same three prevalences of disease, both tests would miss between 10 and 42 melanomas, with the addition of dermoscopy reducing false positives (or reducing the number of excisions that would be performed) by 191, 176 and 158 per 1000 (compared with 238, 220 and 198 unnecessary excisions with visual inspection alone).

We noted very similar findings between the analysis of all studies, and the analyses restricted to studies which made within-person comparisons of strategies of visual inspection alone and visual inspection aided by dermoscopy. The same difference was evident for our secondary analyses for the detection of invasive melanoma alone and for the detection of any skin lesion requiring excision.

2) In-person dermoscopy is substantially more accurate than image-based assessments

Much of the available evidence for the diagnostic accuracy of dermoscopy is based on the interpretation of dermoscopic images (60/86) as opposed to ‘real time’ diagnosis face-to-face with the patient concerned. Formal comparison of test accuracy found in-person dermoscopy to be substantially more accurate compared to diagnosis based on dermoscopic images (RDOR 4.6, 95% CI 2.4, 9.0; P<0.001). Although there may be a number of contributing factors, including differences in study populations, different algorithms to assist test interpretation and differences in observer experience, it is likely that, as for visual inspection of a clinical image (Dinnes 2018a), remote test interpretation cannot approximate a physical, face-to-face patient to clinician interaction. In particular, total body skin examination is likely to have a significant impact on the decision to excise a lesion suspected to be melanoma (Grob 1998; Argenziano 2012; Aldridge 2013). Across the 60 image-based evaluations, half (30/60) were blinded to all other patient information and only 17 (28%) provided observers with the clinical image of the same lesion to assist test interpretation.

Nevertheless, given the increasing trend towards remote test interpretation (or teledermatology) it is important to try to understand the potential impact from image-based assessments. From the data observed, at a fixed specificity of 80%, diagnosis based on dermoscopic images was 34% more sensitive than diagnosis based on clinical images alone (an increase from 47% to 81% sensitivity). Assuming melanoma or intraepidermal melanocytic variant prevalences of 18%, 24% and 39%, these results translate to 164, 152 and 122 false positive test results, with 34, 46 and 74 melanomas missed (false negatives) using dermoscopic images (a reduction of 61, 81 and 133 compared to diagnosis based on clinical images alone). At a fixed sensitivity of 80%, test specificity for diagnosis based on dermoscopic images would be 40% higher compared to that based on clinical images (specificity of 82% compared to 42%). Applying these results to a cohort of 1000 lesions would miss between 36 and 78 melanomas, with 148, 137 and 110 false positive results based on dermoscopic image interpretation (a reduction of 328, 304 and 244 in comparison to the evaluation of clinical images alone).

A *post hoc* analysis restricting study inclusion to those which did not use a case-control design appeared to increase the accuracy of image-based visual inspection and, to a lesser extent, the accuracy of diagnosis based on dermoscopic images. Nevertheless the observed accuracy of in-person dermoscopy was still greater than that using dermoscopic images. It is also important to note that none of the included image-based dermoscopy evaluations purported to be an evaluation of teledermatology. Such evaluations are included in a separate systematic review of teledermatology for the diagnosis of skin cancer (Chuchu 2018a). Although the results for image-based dermoscopy from this review have some bearing on the accuracy that might be achieved by the remote assessment of dermoscopic images, we suggest that future studies should not be undertaken which evaluate dermoscopic images to approximate to in-person evaluation. We have retained the image-based studies in the review as they do enable comparisons of different aspects of dermoscopic diagnosis (see below), but they could potentially be excluded from future reviews.

3) No effect from prior testing of participants or study position on the clinical pathway could be determined, and there is insufficient evidence to assess the accuracy of dermoscopy in a primary care setting.

Less than half of in-person evaluations (42%; 11/26) and only 18% of image-based evaluations (11/60) contributing to analyses for the primary target condition contained enough information to describe the position of participants on the clinical pathway. This figure is lower than for our review of visual inspection for the detection of melanoma, where two thirds of in-person evaluations were clearly positioned on the clinical pathway (Dinnes 2018a). The majority of evaluations of dermoscopy however appear to have been conducted in referral settings, with only four eligible studies conducted in primary care populations; two in-person evaluations and two image-based, thus our planned comparison between initial presentation versus referred patients is underpowered. Within the referred population studies there was some (largely non-significant) indication of higher accuracy in equivocal lesions and lower accuracy in studies of patients with lesions undergoing follow-up, particularly in image-based studies. The classification of study populations was dependent on the terminology used by the study authors and the groupings may not fully reflect differences between study populations.

4) There is no clear evidence that accuracy is improved by the use of any named or published algorithm to assist diagnosis.

The use of a named or published algorithm to assist dermoscopy interpretation (as opposed to no reported algorithm or reported use of pattern analysis) had no significant impact on accuracy either for in-person (RDOR 1.4, 95% CI 0.34, 5.6; P = 0.17) or image-based (RDOR 1.4, 95% CI 0.60, 3.3; P = 0.22) evaluations. This result was supported by subgroup analysis according to algorithm used. Although the vast majority of data comparing algorithms came from image-based evaluations there is no reason to suggest that the relative accuracy of different approaches to diagnosis would vary according to whether the evaluation was image-based as opposed to in-person, even if in absolute terms accuracy is higher for latter group of studies.

In this instance, data could be pooled separately according to algorithm and threshold used therefore the bivariate normal model was employed rather than the summary ROC approach. For in-person evaluations most of the data related to no algorithm (8 datasets), to pattern analysis (6 datasets) or to the ABCD approach at a threshold of >5.45 (5 datasets). Test sensitivities and specificities were broadly similar for no algorithm (88%, 95% CI 75, 95% and 87%, 95% CI 80, 92%) and for pattern analysis (92%, 95% CI 87, 95% and 92%, 95% CI 68, 98%); use of the ABCD algorithm produced similar specificity (92%, 95% CI 82, 97%) but lower sensitivity (81%, 95% CI 62, 92%), although confidence intervals were wide and overlapping. At the median prevalence of melanoma of 12% observed across the in-person evaluations, the number of melanomas missed per 1000 lesions tested ranged between 10 and 23 with false positives of 70 to 114 (Summary of findings table 1). For image-based evaluations, test sensitivities and specificities were again broadly similar for no algorithm (76%, 95% CI 70, 82% and 79%, 95% CI 71, 85%) and for pattern analysis (83%, 95% CI 76, 88% and 87%, 95% CI 80, 92%). The formal algorithms with the most data included ABCD at >5.45 (7 datasets), the seven point checklist at >=3 (11 datasets) and the three point checklist (7 datasets), sensitivities were broadly similar with overlapping confidence intervals (ranging from 74% to 81%), with generally lower specificities but again with overlapping confidence intervals (summary estimates ranging from 60% to 81%). At the median prevalence of melanoma of 24% observed across the image-based evaluations, the number of melanomas missed per 1000 lesions tested ranged between 41 and 62 with false positives of 61 for no algorithm to 304 for the three point checklist (Summary of findings table 1).

The lack of reporting of diagnostic thresholds in the studies that did not use algorithms to assist diagnosis ('no algorithm' studies) means that we have not been able to clearly compare accuracy for the *diagnosis* of melanoma in comparison to a clinician's decision to *excise* a skin lesion; the latter perhaps being more clinically relevant in practice. Data from image-based studies appears to show similar sensitivity for correct diagnosis of melanoma and for the decision to excise a lesion but considerably lower specificity for the decision to excise a lesion, when the target condition was defined as melanoma or atypical intraepidermal melanocytic variants. For the target condition of any skin cancer or lesion with a high risk of progression to melanoma, sensitivities and specificities were both over 90% in three of the four studies reporting data for dermoscopy added to in-person visual inspection, suggesting that clinicians may be better at identifying skin lesions that require some intervention than at correctly identifying melanomas, however the data are too limited to allow strong conclusions to be drawn.

5) Observer expertise and training in dermoscopy improves diagnostic accuracy

Observer experience and expertise in using dermoscopy to assess pigmented lesions is likely to impact on test accuracy, however this information was often not provided in great detail, particularly for the in-person evaluations. Broad classifications of reported experience in dermoscopy and by observer qualifications were made which, on the whole, led to statistically significantly higher accuracy for observers reported as having high experience and for those classed as ‘expert consultants’ in comparison to those considered to have less experience in dermoscopy. Much of the evidence for the effect of observer expertise was again provided by image-based dermoscopy interpretations as opposed to those conducted in-person, however similar patterns were observed for both sets of studies. Only 2 in-person and 3 image-based studies evaluated dermoscopy in the hands of GPs; these showed lower accuracy (RDOR 0.21 (95% CI 0.01, 3.12) for in-person and RDOR 0.09 (95% CI 0.04, 0.24) for image-based studies) than expert consultants.

Six studies assessed the effect of dermoscopy training on test accuracy in a limited number of participants. Despite differences in the type and length of training interventions, all of the six eligible evaluations resulted in increased sensitivity following training with limited effects on specificity in five of the six studies.

Strengths and weaknesses of the review

The strengths of this review include an in-depth and comprehensive electronic literature search, systematic review methods including double extraction of papers by both clinicians and methodologists, and contact with authors to allow study inclusion or clarify data. A clear analysis structure focusing on estimating incremental gains in accuracy was adopted. A detailed and replicable analysis of methodologic quality was undertaken.

For our main analyses however, summary ROC curves were estimated rather than average sensitivity and specificity operating points. This approach was undertaken to facilitate pooling across the heterogeneous mixture of thresholds and scoring systems, however it does mean that quoted sensitivities and specificities are at best illustrative and do not reflect the actual performance of dermoscopy. As a result, although we can assess the incremental gain from dermoscopy added to visual inspection, we cannot make any clear recommendations regarding how dermoscopy should be used in order to ensure that melanomas are not missed.

In comparison to other available systematic reviews, our review extends the time period searched for eligible studies, and includes all eligible studies regardless of availability of a direct comparison with visual inspection alone (Vestergaard 2008), requirement for an algorithm or 'clinical prediction rule' (Harrington 2017), or focus on specific health care professionals or study settings (Corbo 2012; Loescher 2011; Herschorn 2012). Our review of a single large literature search and concurrent systematic review of a number of other tests for the diagnosis of melanoma has led to the identification of additional dermoscopy datasets and inclusion of a much greater of number of studies (i.e. 104 compared to 23 in Rajpara 2009; 9 in Vestergaard 2008; and 43 in Harrington 2017). We also explicitly considered whether diagnoses were made based on dermoscopic images or were conducted in-person and considered variations in the definition of the target condition. Most importantly perhaps our review considers the accuracy of dermoscopy both in comparison to visual inspection and for diagnosis with and without the use of a formal algorithm. As for considerations of the accuracy of visual inspection of a lesion *per se* (Dinnes 2018a), unless the accuracy of diagnostic decisions made without the use of a formal algorithm can be established, the added contribution of such algorithms cannot be fully understood.

Our stringent application of review inclusion criteria meant that some studies included in previous reviews were excluded. For example, those reporting accuracy data for 'clinical diagnosis' where dermoscopy may or may not have been used to assist diagnosis were not included. Of the nine studies included in the Vestergaard 2008 review, we excluded two due to the inclusion of less than five melanomas (Carli 2003a; Carli 2004) and of the 23 in Rajpara 2009 we excluded one due lack of clarity on the 2x2 contingency table (Ascierto 2000). Seven of the 43 studies included in the Harrington review were also excluded due to lack of clear data to construct a 2x2 contingency table (Argenziano 2003) or reporting of data in brief letter format (Blum 2004a; Strumia 2003), the serial use of the algorithm in the context of lesion follow-up (Buhl 2012), the derivation aspect to the study (Henning 2008; MacKie 2002) or diagnosis by laypersons (Luttrell 2012).

The main concerns for the review are a result of the poor reporting of primary studies, in particular limiting assessment of methodological quality, and limiting both the assessment of studies by prior testing of participants and by observer expertise in dermoscopy. Our review of visual inspection alone for the diagnosis of melanoma identified a general trade-off between sensitivity and specificity along the clinical pathway with higher sensitivity and lower specificity in limited prior testing studies compared to those in referred populations (Dinnes 2018a). The lack of data from limited prior testing populations in this review and the lack of detailed information on the prior testing of participants included in referred populations meant that no clear patterns in sensitivity or specificity could be derived. Some evidence of higher accuracy by more specialist or experienced observers was identified however better study descriptions of observers would assist such investigations.

Applicability of findings to the review question

There are clear concerns regarding the clinical applicability of studies included in this review. Approximately three quarters of studies only provided data from evaluations of dermoscopic images (with or without data from visual inspection of clinical photographs) such that resulting accuracy estimates cannot be extrapolated to in-person assessments of skin lesions. Furthermore, almost all in-person evaluations of dermoscopy used in conjunction with visual inspection had high concerns for the applicability of the included population and half had high concern for the applicability of the test. The restriction of including only excised lesions and the small number of studies conducted in a limited prior testing population mean that our results cannot be extrapolated to a primary care population.

Authors' conclusions

Implications for practice

Due to methodological limitations of the included studies and heterogeneity in study methods and results, the sensitivity and specificity of dermoscopy either with or without visual inspection cannot be explicitly estimated, however, we can conclude that the incremental benefit of dermoscopy over and above visual inspection alone is consistent and considerable. Dermoscopy is therefore a valuable tool to support visual inspection of a suspicious skin lesion for the detection of melanoma and atypical intraepidermal melanocytic variants, particularly in referred populations and in the hands of experienced users. Data to support its use in a primary care population is limited; however, it is likely to be of some benefit for triaging suspicious lesions for urgent referral when employed by suitably trained clinicians. Overall, the use of formal algorithms to assist diagnosis does not appear to improve accuracy, however, neither is there sufficient evidence to suggest that the 'no algorithm' approach should be preferred in all settings. Formal algorithms may be more useful for dermoscopy training purposes and for less expert observers, however reliable data from in-person evaluations of dermoscopy are lacking.

Implications for research

Given the vast volume of research that has been funded to evaluate dermoscopy, further research into the added value of established dermoscopy algorithms *per se* is unlikely to be warranted. Further evaluation of dermoscopy use in the primary care setting and to identify the optimal approach to dermoscopy training may be warranted, however. Such evaluations should be conducted on an in-person basis with prospective recruitment of consecutive series of participants and with systematic follow-up of non-excised lesions to avoid over-reliance on a histological reference standard. A clear identification of the level of training and experience required to achieve good results is required. Any future research study needs to be clear about the diagnostic pathway followed by study participants prior to study enrolment, and should conform to the updated Standards for Reporting of Diagnostic Accuracy (STARD) guideline (Bossuyt 2015).

Acknowledgements

Members of the Cochrane Skin Cancer Diagnostic Test Accuracy Group include:

- the full project team (Susan Bayliss, Naomi Chuchu, Clare Davenport, Jonathan Deeks, Jacqueline Dinnes, Lavinia Ferrante di Ruffano, Kathie Godfrey, Rubeta Matin, Colette O'Sullivan, Yemisi Takwoingi, Hywel Williams)
- our 12 clinical reviewers (Rachel Abbott, Ben Aldridge, Oliver Bassett, Sue Anne Chan, Alana Durack, Monica Fawzy, Abha Gulati, Jacqui Moreau, Lopa Patel, Daniel Saleh, David Thompson, Kai Yuen Wong) and 2 methodologists (Lavinia Ferrante di Ruffano and Louise Johnston) who assisted with full text screening, data extraction and quality assessment across the entire suite of reviews of diagnosis and staging and skin cancer,
- our expert advisor and co-author Fiona Walter
- and all members of our Advisory Group (Jonathan Bowling, Colin Fleming, Matthew Gardiner, Abhilash Jain, Susan O'Connell, Pat Lawton, John Lear, Mariska Leeflang, Richard Motley, Paul Nathan, Julia Newton-Bishop, Miranda Payne, Rachael Robinson, Simon Rodwell, Julia Schofield, Neil Shroff, Hamid Tehrani, Zoe Traill, Fiona Walter).

Cochrane Skin editorial base wishes to thank Michael Bigby, who was the Dermatology Editor for this review; and the clinical referee, Adam Bray. We also wish to thank the Cochrane DTA editorial base and colleagues.

Contributions of authors

JD was the contact person with the editorial base.
JD co-ordinated contributions from the co-authors and wrote the final draft of the review.
JD, NC, LFR, DT, KYW, RBA, RA, and MF screened papers against eligibility criteria.
JD and NC obtained data on ongoing and unpublished studies.
JD, NC, LFR, DT, KYW, RBA, RA, and MF appraised the quality of papers.
JD, NC, LFR, DT, KYW, RBA, RA, and MF extracted data for the review and sought additional information about papers.
JD entered data into RevMan.
JD, MJG and JJD analysed and interpreted data.
JD, JJD, NC, LFR, YT and CD worked on the methods sections.
JD, FW, DT, KYW, RBA, RA, MF, RNM and HCW drafted the clinical sections of the background and responded to the clinical comments of the referees.
JD, JJD, CD and YT responded to the methodology and statistics comments of the referees.
KG was the consumer co-author and checked the review for readability and clarity, as well as ensuring outcomes are relevant to consumers.
JD is the guarantor of the update.

Disclaimer

This project was supported by the National Institute for Health Research, via Cochrane Infrastructure funding to the Cochrane Skin Group and Cochrane Programme Grant funding. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, NHS or the Department of Health.

Declarations of interest

Jacqueline Dinnes: I am employed by the University of Birmingham under a National Institute for Health Research (NIHR) Cochrane Programme Grant (13-89-15) to produce this review.
Jonathan J Deeks: The project was funded by the NIHR Cochrane Programme Grant. Jon Deeks received additional support as an NIHR Senior Investigator and from the NIHR Birmingham Inflammation Biomedical Research Centre.
Matthew J Grainge: nothing to declare.
Naomi Chuchu: nothing to declare.
Lavinia Ferrante di Ruffano: The NIHR Cochrane Systematic Review Programme Grant funded the project but to my knowledge did not influence its planning or conduct.
Rubeta N Matin: nothing to declare.
David R Thomson: nothing to declare.
Kai Yuen Wong: nothing to declare.
Roger Benjamin Aldridge: nothing to declare.

Rachel Abbott: nothing to declare.
Monica Fawzy: nothing to declare.
Susan E Bayliss: nothing to declare.
Yemisi Takwoingi: nothing to declare.
Clare Davenport: My employer (The University of Birmingham) received funding for my participation in this review as part of an NIHR clinical fellowship awarded to Alex Heazell (the lead author and contact person).
Kathie Godfrey: I have received reimbursement of travel expenses incurred by attending meetings.
Fiona M Walter: nothing to declare.
Hywel C Williams: I am director of the NIHR HTA Programme. HTA is part of the NIHR which also supports the NIHR systematic reviews programme from which this work is funded.

Differences between protocol and review

We set out to review visual inspection and dermoscopy for the detection of melanoma in a single review, however due to the volume of evidence identified, two separate reviews were prepared: one for visual inspection alone and one for dermoscopy. This review of dermoscopy includes data for the accuracy of visual inspection but only where both tests were evaluated in the same studies (direct comparisons).

Primary objectives and primary target condition have been changed from detection of cutaneous invasive melanoma alone, to the detection of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants, as the latter is more clinically relevant to the practicing clinician. The detection of the target condition of invasive melanoma alone has instead been included as a secondary objective. The primary objectives were also amended to conduct separate analyses by in-person/image-based diagnosis rather than to investigate the effect on accuracy as a secondary objective, as originally proposed in the generic protocol. This decision was taken very early in the review process and was based on the fact that a diagnosis based on a dermoscopic image or clinical photograph cannot approximate the context of a face-to-face patient clinician consultation.

Secondary objectives have been tailored to the individual test, with two objectives added: to determine the diagnostic accuracy of individual algorithms for dermoscopy; to determine the effect of observer experience; and to determine the effect on accuracy of observer training in dermoscopy.

Sources of heterogeneity that could be investigated (as listed in the protocol) were restricted due to lack of data.

Studies using cross-validation, such as 'leave-one-out' cross-validation were *excluded* rather than included as these methods are not sufficiently robust and are likely to produce unrealistic estimates of test accuracy.

To improve clarity of methods, this text from the protocol "We will include studies developing new algorithms or methods of diagnosis (i.e., derivation studies) if they use a separate independent 'test set' of participants or images to evaluate the new approach.We will also include studies using other forms of cross validation, such as 'leave-one-out' cross-validation (Efron 1983). We will note for future reference (but not extract) any data on the accuracy of lesion characteristics individually, e.g., the presence or absence of a pigment network or detection of asymmetry" has been replaced with the following: "Studies developing new algorithms or methods of diagnosis (i.e., derivation studies) were included if they:

- used a separate independent 'test set' of participants or images to evaluate the new approach, or
- investigated lesion characteristics that had previously been suggested as associated with melanoma and the study reported accuracy based on the presence or absence of particular combinations of characteristics.

Studies were excluded if they:

- used a statistical model to produce a data driven equation, or algorithm based on multiple diagnostic features, with no separate test set.
- used cross-validation approaches such as 'leave-one-out' cross-validation (Efron 1983)
- evaluated the accuracy of the presence or absence of individual lesion characteristics or morphological features, with no overall diagnosis of malignancy
- reported accuracy data for 'clinical diagnosis' with no clear description as to whether the reported data related to visual inspection alone or included dermoscopy in all study participants·
- were based on the experience of a skin cancer-specific clinic, where dermoscopy may or may not have been used on an individual patient basis."

We proposed to supplement the database searches by searching the annual meetings of appropriate organisations (e.g., British Association of Dermatologists Annual Meeting, American Academy of Dermatology Annual Meeting, European Academy of Dermatology and Venereology Meeting, Society for Melanoma Research Congress, World Congress of Dermatology, European Association of Dermato Oncology), however due to volume of evidence retrieved from database searches and time restrictions we were unable to do this.

For quality assessment, the QUADAS-2 tool was further tailored according to the review topic. In terms of analysis, restriction to analysis of per patient data was not performed due to lack of data.

Published notes

Characteristics of studies

Characteristics of included studies

Ahnlide 2016

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective **Period of data collection** 7 March 2013 to 28 April 2014 **Country** Sweden

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Excised melanocytic skin lesions with recorded dermoscopy ABCD score and clinician's preliminary diagnosis. **Setting:** Secondary (general dermatology) **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** Previously biopsied lesions and wide excisions not included; other exclusion prior to enrolment included: invalid report or missing data (n=34); visiting residents data (n=66); non-melanocytic on histology or benign melanocytic lesions with special patterns (e.g. papillomatous, congenital naevi and mucosal lesions) (n=658) **Sample size (patients):** NR **Sample size (lesions):** No. eligible: 1135/ No. included: 309 **Participant characteristics:** NR **Lesion characteristics:** NR

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm (Clinician's preliminary diagnosis); ABCD **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** ABCD >4.75 or >5.45 (calculated automatically based on clinician scoring presence/absence of ABCD criteria into computerized patient file) Preliminary preoperative diagnosis was based on physical examination and dermoscopic assessment (including application of ABCD algorithm) **Diagnosis based on:** Single observer (n=13);

	Observer qualifications: Dermatology residents (n=6; "residents were encouraged to consult the specialists in difficult cases"); Dermatologists (n=7) **Experience in practice:** Not described **Experience with dermoscopy:** Not reported per observer, but assumed High given training described; describe use of dermoscopy and ABCD at department for more than 10 years; reports "repeated joint feedback sessions evaluating the preoperative dermoscopy photographs of excised lesions, enrolment in dermoscopy courses for both residents and senior consultants and daily continuous education in dermoscopy for residents."

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	No
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** - Histological diagnosis alone; histopathological diagnosis was recorded postoperatively in the patient file by a nurse. Disease positive: 46; Disease negative: 263 **Target condition (Final diagnoses)** Melanoma (invasive): 23; Melanoma (in situ): 23 Benign naevus: 263
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: Missing scoring (n=57); wrongly scored due to pre-op non-melanocytic diagnosis (n=5); lesions with preliminary diagnosis of lentigo maligna or Spitz naevus (n=5); ambiguous histology (n=1). **Time interval to reference test**: Not reported-but likely consecutively as dermoscopy was used pre-operatively
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Alarcon 2014

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective; dermoscopic images assessed remotely from the patient **Period of data collection** 1 June 2011 and 30 May 2012 -1 year **Country** Spain
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Dermoscopically equivocal pigmented lesions, assumed to be melanocytic, seen at Melanoma Unit **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) Melanoma Unit of the Hospital Clinic of Barcelona. **Prior testing:** Dermatoscopic suspicion in all cases **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** Non-melanocytic appearance **Sample size (patients):** No. eligible: unclear/ No. included: 264 **Sample size (lesions):** No. eligible: 343/ No. included: 264 **Participant characteristics:** Median age (yrs)**:** 54.7 (8-89y); 51.5% Male; **Lesion characteristics:** Fitzpatrick phototype: I to II - 42%; III to IV 50%; Lesion site: Head/Neck: 73; 27.7%; Trunk: 135; 51.1%; Limbs: 49; 18.6%; Describe if other 7; 7% (acral). Lesion thickness: ≤1mm: 86 of 92 melanoma
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** Clinical examination and/or case notes - lesion site and age provided plus RCM images. Dermoscopy and RCM interpretation appear to have been conducted by same observer with no indication of blinding **Diagnostic threshold:** Not reported; no details **Diagnosis based on:** Single observer (n=3) **Observer qualifications:** Dermatologist. All the images were interpreted independently by one of the three dermatologists with expertise in RCM **Experience in practice:** Not described **Experience with dermoscopy:** Assumed High experience; three dermatologists with expertise in RCM **Any other detail** All of the lesions were imaged with a digital camera (Canon PowerShot G10; Canon, Tokyo, Japan) and a high-resolution dermatoscope dermatoscope (DermLite Photo; 3Gen LLC, Dana Point, CA, U.S.A.).

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus follow-up. Histology (n=264); Follow-up (n=79); selection for excision based on RCM diagnosis otherwise all would have been excised **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 92; BCC: 12 Benign naevus: 107; 53 SK and AK
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** none reported **Time interval for reference:** Appears consecutive; "Data regarding age, sex, anatomical location, melanoma risk factors and dermoscopic diagnosis were collected before the RCM examination and histopathological analyses were performed" **Time interval between index test(s):** Not specified but appears consecutive application of dermoscopy and RCM
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes

If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Annessi 2007

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Not reported **Period of data collection** Dec 2004--June 2006 **Country** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Consecutive atypical macular melanocytic lesions; all larger than 5 mm in diameter, with a flat or barely elevated surface and at least 3 of the following features: (a) asymmetry, (b) irregular margins, (c) ill-defined borders, and (d) color variegation. **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** None reported **Sample size (patients):** No. included: 195 **Sample size (lesions):** No. included: 198 **Participant characteristics:** Mean age: 43 years; Male: (106 males) 54% **Lesion characteristics:** all <=1mm thickness; mean 0.3mm; all >5mm diameter

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy.** Pattern analysis; 7-point checklist ; ABCD **Method of diagnosis:** Dermoscopic images **Prior test data:** Unclear; clinical and ELM digital images taken but unclear what was actually presented to observers **Diagnostic threshold:** Reported only for ABCD - Melanocytic lesions with ABCD scores between 4.76 and 5.45 (suspect lesions) were considered test positive **Diagnosis based on:** Consensus (n=2) **Observer qualifications:** Described as "ELM-experienced dermatologists" **Experience in practice:** High experience or 'Expert' **Experience with dermoscopy:** High experience /'Expert' users

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias

Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone; conducted in Dermatopathology laboratory Disease positive: 96; Disease negative: 102 **Target condition (Final diagnoses)** Melanoma (invasive): 72 ; Melanoma (in situ): 24 Benign naevi: 102 - described as Clark's melanocytic nevi (68 junctional and 34 compound)

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** None described **Time interval to reference test:** Appears consecutive; "After ELM assessment, all lesions were excised and processed for routine histopathologic examination"

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Unclear
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Arevalo 2008

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** No time period given just states lesions evaluated since 1991 **Country** Australia

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic lesions imaged at the Sydney Melanoma Unit with a histopathologic diagnosis or that remained unchanged following short-term (5-4.5 months) digital monitoring (diagnosed as benign) **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Selected for excision (no further detail); Changes on digital monitoring **Setting for prior testing:** Not reported **Exclusion criteria:** Lentigo maligna and lentigo malignant melanoma **Sample size (patients):** NR **Sample size (lesions):** No. eligible: 3367 melanocytic lesions/ No. included: 3367 **Participant characteristics:** NR **Lesion characteristics** NR

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** Menzies criteria **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used **Diagnostic threshold:** lesion must have none of the 2 negative features of symmetry of pattern or single colour, and must have 1 or more of the following 9 positive features of melanoma; blue-white veil, pseudopods, radial streaming, peripheral black dots or globules, multiple brown dots, multiple blue-gray dots, scar like depigmentation, broadened network and multiple colours. **Diagnosis based on:** Unclear; appears to be consensus (n=2); all lesions scored independently by 2 observers blinded to the diagnosis, with referral to a third observer if there was a disagreement. **Observer qualifications:** Not reported; likely dermatologists **Experience in practice:** Not described **Experience with dermoscopy:** Not described **Any other detail** The images were obtained using a dermoscopic camera (Dermaphot; Heine Ltd) or a digital imaging device (Solarscan).

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** - Histological diagnosis plus follow up Further details: Not described in detail; Only included lesions with histopathology or those that remained unchanged following short-term (2.5-4.5 months) **Target condition (Final diagnoses)** Melanoma (invasive): 341 'Benign' diagnoses: 3026
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: Poor quality index test image as exclusion criterion **Time interval to reference test:** Not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	No
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Argenziano 1998

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Unclear **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** Not reported **Country:** Italy **Test set derived** Three hundred forty two lesions were randomly divided into a training set of 57 CMs and 139MN and a test set of 60 CMs and 86MN.
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Unclear
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Atypical melanocytic skin lesions with dermoscopic images that had undergone biopsy due to clinician suspicion **Setting:** Not reported **Prior testing:** Dermatoscopic suspicion in all cases **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** None reported **Sample size (patients):** NR

	Sample size (lesions): No. eligible: 342/ No. included: 342 **Participant characteristics:** NR **Lesion characteristics:** Lesion thickness - ≤1mm: 28%; 68 CMs <0.76 mm; 49 CMs >0.75 mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Pattern analysis (Not described but classified by clinical reviewer); 7-point checklist (derived and evaluated in this study); ABCD **Method of diagnosis:** Dermoscopic images; in vivo photography as x10 magnification with special photography equipment after being covered in immersion oil. **Prior test data:** No further information used; "In a blind study"- implies no information beyond the dermoscopic images available. **Diagnostic threshold:** Pattern analysis - 'overall ELM diagnosis; 'ABCD - Score >4.75; 7-point checklist - Score of 3 or more. **Diagnosis based on:** Single (n=2; less experienced observers) and Consensus (2 observers) (n=3; ELM-experienced observers) **Observer qualifications:** Dermatologist **Experience in practice:** Not described. **Experience with dermoscopy:** High experience - 3 ELM-experienced; Moderate/trained - less experienced dermatologists (who underwent "short formal ELM training of 9 hours") **Any other detail:** Training set used to derive 7-point checklist. Initially two models were developed. One using multivariate analysis to create a formula for calculating the probability of each lesion belonging to the group of melanomas but was deemed too complex for clinical use. The second model used the odds ratios (ORs) from the multivariate analysis to create a simpler diagnostic method based on identification of major and minor ELM criteria. A score of 2 was given to the 3 criteria with ORs>5 (major criteria) and a score of 1 was given to the 4 criteria with OR< 5 (minor criteria); a total score of 3 or more set to id melanoma. Major criteria included atypical pigment network (presence of an irregular and prominent pigment network), gray-blue areas and atypical vascular pattern. Minor criteria: streaks, blotches, irregular dots and globules, and regression pattern (presence of white areas or peppering).

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described) Disease positive: 117; Disease negative: 225 **Target condition (Final diagnoses)** Melanoma (invasive): 99; Melanoma (in situ): 18 'Benign' diagnoses: 114 atypical nevi111 common nevi
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** None reported **Time interval to reference test:** Not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias

Comparative	

B. Concerns regarding applicability

Notes

Notes	

Argenziano 2006

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Randomised controlled trial allocating primary care physicians to use either visual inspection alone or visual inspection plus dermoscopy (only excised lesions can be included for each arm) **Data collection:** Prospective **Period of data collection** May 2003 to Sept 2004 **Country** Italy and Spain

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients asking for screening or exhibiting one or more skin tumours as seen during routine physical examination (patient-finding screening) were considered for inclusion; those undergoing excision were included in this review (i.e. those deemed sufficiently suspicious by the Expert evaluation). PCPs were invited to participate in the trial; only those who attended the training sessions and who then screened patients and referred them to the Pigmented Lesion Clinics were randomised. **Setting:** Primary **Prior testing:** No prior testing **Setting for prior testing:** N/A **Exclusion criteria:** NR **Sample size (patients):** No. eligible: 3271 patients screened; 1325 patients allocated to Naked Eye observation and 1197 patients allocated to dermoscopy observation; No. included: 162 received histology after Expert evaluation at the PLC **Sample size (lesions):** 85 in VI arm and 77 in Dermoscopy arm underwent excision **Participant characteristics:** Based on full sample: mean age 40, range 2-90 (visual inspection group)/ 41, range 3-94 (dermoscopy group). Male 498 (38%) : VI group / 451 (38%) dermoscopy **Lesion characteristics** NR

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** ABCD (control arm of RCT) **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Diagnostic threshold:** Qualitative NR; Described in Intro as: simple morphologic features summarized by the asymmetry, border irregularity, color variegation, and diameter 5 mm (ABCD) **Diagnosis based on:** Average (n=37) **Observer qualifications:** Primary care physicians **Experience in practice:** Not described **Experience with dermoscopy:** Not described **Other detail:** Pre-randomisation all participating PCPs underwent training in ABCD rule for clinical diagnosis and 3-point checklist for dermoscopy (see below). **Dermoscopy** 3-point rule (intervention arm of RCT) **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Diagnostic threshold:** >=2 chars present (algorithm is based on the recognition of only three individual features: dermoscopic asymmetry (in color and/or structure, not in shape), atypical network (pigmented network with thick lines and irregular distribution), and blue-white structures (presence of any blue and/or white color within the lesion). Each PCP in both groups examined the individual lesions and scored the patient outcome, as banal or suggestive of skin cancer **Diagnosis based on:** Average (n=36) **Observer qualifications:** Primary care physicians **Experience in practice:** Not described **Experience with dermoscopy:** Not described **Dermoscopy training:** All PCPs received training (2 hour session) on the clinical ABCD rule for diagnosis of melanoma, basic recognition of nonmelanoma skin cancers including BCC and SCC plus a 2 hour session describing the dermoscopy 3-point checklist.

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone All lesions considered suggestive of skin cancer at the PLC were excised and subsequently diagnosed histopathologically. Equivocal lesions by histopathologic examination were reviewed by a second independent pathologist and a final diagnosis made. Disease positive: 92 malignant tumours; Disease negative: 70 benign tumours **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 12; BCC: 66; cSCC: 14 Sebhorrheic keratosis: 13; Melanocytic nevi = 51; Other: 6
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** only those patients who were considered to have lesions suggestive of skin cancer had histology and were included. All the rest had expert diagnosis (not included in the final 2x2 data extracted) **Time interval to reference test:** Not reported **Time interval between index test(s)**: N/A (RCT)
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	RCT examining effect of making dermoscopy available to primary care practitioners
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	Yes
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	Yes
Are there concerns that the test comparison differs from the review question?	Low concern

Notes

Notes	

Argenziano 2011

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective **Period of data collection:** 2006-2008 **Country:** Naples, Italy
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Randomly sampled 100 melanomas and 100 excised melanocytic naevi from a digital collection of lesions screened between 2006 and 2008 at the Department of Dermatology of the Second University of Naples; also randomly sampled 100 melanocytic naevi that showed no relevant changes to warrant excision during the follow up period from a larger database of monitored naevi

	Setting: Secondary (general dermatology) **Prior testing:** Retrospective study of a random sample of dermoscopic images collected in departmental database.100/349 excised melanomas 100/1512 excised naevi **Setting for prior testing:** Not reported **Exclusion criteria:** Excluded non-melanocytic lesions, lesions on certain anatomical sites (facial, acral, mucosal and nail lesions), lesions larger than 15 mm, and lesions with conflicting histopathological features **Sample size (patients):** NR **Sample size (lesions):** No. included: 300 **Participant characteristics:** NR **Lesion characteristics:** NR

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Pattern analysis; 7-point checklist; revised 7-point checklist **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used; "No additional information was provided, to avoid the possible bias that clinical information may give to the assessment on morphological criteria." **Diagnostic threshold:** Pattern analysis - classify as naevus/melanoma/ or lesion to be excised. 7-point checklist - individual criteria scored. Original 7-point - score >=3 merits excision (based on three major criteria with 2 points each (atypical network, blue-white veil and atypical vascular pattern) and four minor criteria with 1 point each (irregular dots/globules, irregular streaks, irregular blotches and regression structures). Revised 7-point checklist: score>=1 merits excision (each criterion is given a score of 1 point). **Diagnosis based on:** Average; (n=8) **Observer qualifications:** Dermatologist **Experience in practice:** High; 'Experienced dermatologists' **Experience with dermoscopy:** High; Dermatologists specifically trained in dermoscopy

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus follow up; 200/300 had histology. 100/300 were naevi that had been followed up 1-3 years (median 22 months; range 1-3 years). **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 100; not clear if *in situ* included. Excised naevi included: 57 Clark naevi, 28 Spitz naevi, 10 small congenital naevi and 5 blue naevi. The remaining 100 monitored lesions were reported as 74 reticular naevi and 26 globular naevi;
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** None reported **Time interval to reference test:** Unknown
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Yes

If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Unclear
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Ascierto 2010

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** Not reported (states in a period of 1 year) **Country** Italy

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Clinically relevant cutaneous pigmented lesions, undergoing dermoscopy and excision; only melanocytic lesions meeting at least two clinical ABCDE criteria underwent dermoscopy **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Clinical examination with ABCDE **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** None reported **Sample size (patients):** No. eligible: 54/ No. included: 54 **Sample size (lesions):** NR **Participant characteristics:** Median age 41 (19-73y); 19 males **Lesion characteristics** NR

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Risk stratification (modified Kenet et al) **Method of diagnosis:** In person diagnosis; all patients underwent total body skin examination **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** Very high risk - Lesion with a pigment network and any of the classical ELM features specific for melanoma (pseudopods, radial streaming, blue-gray veil, atypical vessel, etc.). High risk - Lesion with a pigment network and subtle new ELM features that may suggest melanoma but often are also seen in atypical nevi. **Diagnosis based on:** Unclear, assumed Single observer per pt (n=3) **Observer qualifications:** Dermatologist **Experience in practice:** High; "evaluations made by expert dermatologists (at least 3 years of experience)" **Experience with dermoscopy:** Assumed High

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described)

	Disease positive: 12 MM; Disease negative: 42 **TARGET CONDITION (Final diagnoses)** Melanoma (invasive) 12 'Benign' diagnoses: 42

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Exclusions:** none reported **Time interval to reference test**: " Before surgery, all patients were investigated by clinical and epiluminescence microscopy (ELM) screenings;"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Bauer 2000

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Not reported. Appears retrospective. **Period of data collection** January 1996 to February 1997 **Country** Italy
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions examined and excised during a campaign for the early diagnosis of cutaneous melanoma (CM) **Setting:** Secondary (general dermatology); From authors' institution **Prior testing:** Not reported "campaign for the early diagnosis of cutaneous melanoma (CM)" **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** No. included: 311 **Sample size (lesions):** No. included: 315 **Participant characteristics:** NR **Lesion characteristics:** Thickness: 14 <0.75 mm, 10 0.75 to 1.5 mm, and 6 >1.5 mm (n=42 melanoma)
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy.** No algorithm; possibly based on Pattern analysis **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination based on ABCD **Diagnostic threshold:** Presence of malignancy; ELM parameters considered included irregular and multi component pigmentary network pattern, peripheral dark network patches, sharp network margin, pseudopods, radial streaming, blue-grey areas, pigment dots (blotches, black dots, brown globules), black dots at periphery, whitish veil, depigmentation and hypopigmented areas, erythema, telangiectasia, comedo-like openings, milia-like cysts, red-blue areas. **Diagnosis based on:** Consensus (3 observers) "diagnosis was made by consensus amongst the dermatologists (Stanganelli 2005) ... when they disagreed a fourth dermatologist, an expert in the diagnosis of PSLs, was consulted."; n=4 **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Assumed High - all dermatologists were "trained in the recognition of PSLs during a training course on the clinical diagnosis of naevi and melanomas"; with referral of disagreements to PSL expert

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described) Disease positive: 42; Disease negative: 273 **TARGET CONDITION (Final diagnoses)** Melanoma (invasive): 30; Melanoma (in situ): 12 Severe dysplasia: 25 'atypical' dysplastic; Benign naevus: 212; 36 nonmelanocytic
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Particpant exclusions:** None reported **Index test to reference standard interval:** After diagnosis, "all lesions were then excised and examined histologically
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Benelli 1999

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** 01/09/1997 to 30/09/1998 **Country** Italy
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** All pigmented skin lesions observed and excised at the Dermatologic Surgery Department **Setting:** Dermatologic Surgery Department **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Dermatologic Surgery Department **Exclusion criteria:** None reported **Sample size (patients):** NR **Sample size (lesions):** No. included: 401 **Participant characteristics:** NR

	Lesion characteristics: Thickness 42 < 0.75 mm thick, 80.76-1.5 mm thick. 4 1.5-4 mm thick (mean 0.60 mm, median 0.55 mm. max 1.9 mm, min 0.10 mm, SD 0.45).

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** ABCDE **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Diagnostic threshold:** Data given for accuracy of each potential score (1-5); score estimation described in detail **Diagnosis based on:** Consensus (2 observers) n= 2 **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Not described **#** **Dermoscopy** 7FFM **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical and dermoscopic evaluations made in-person by 2 dermatologists prior to excision. Decision to excise the lesions was take prior to this by 3 different dermatologists. **Diagnostic threshold:** 2x2 available for 77FM on its own, and for 77FM + each of 5 clinical features, and also for 77FM + each of 5 clinical scores (1-5).; score estimation described in detail **Test observers** as described for Visual Inspection (above)

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	No
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	High risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 60 (15%) lesions; Disease negative: 340 (non melanoma) + 1 BCC **Target condition (Final diagnoses)** Melanoma (invasive): 54 (13.5%); Melanoma (in situ): 6 (1.5%); BCC: 1 (0.4%) Sebhorrheic keratosis: 1 (0.4%); Melanocytic nevi: 316; Epithelioid and/or spindle cell nevi: 18 (4.5%); Lentigo simplex: 5 (1.2%)
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: NR; **Time interval to reference test**: same day

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	**Blinding between tests**: Clinical and dermoscopic evaluations made in-person by 2 dermatologists prior to excision. **Time interval between index test(s)**: same day

Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Benelli 2000

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case-control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** Jan 1993 to Dec 1998 (melanomas); Sep 1997 to Sept 1999 (melanocytic nevi) **Country:** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** All small (<= 6 mm) melanomas and melanocytic nevi consecutively excised over two different time periods **Setting:** Secondary (general dermatology) **Prior testing:** NR; all excised **Setting for prior testing:** NR **Exclusion criteria:** Size > 6mm **Sample size (patients):** NR **Sample size (lesions):** 600 **Participant characteristics:** Mean age 44y (range 20-79) **Lesion characteristics:** NR

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** ABCDE **Method of diagnosis:** Image-based **Prior test data:** Unclear whether dermoscopic image also shown at same time **Diagnostic threshold:** >=2 chars present **Diagnosis based on:** Consensus of 3 (evaluated by 3 different observers; in case of disagreement , the majority view prevailed) **Observer qualifications:** Dermatologist (assumed from authors institution) **Experience in practice:** NR **Experience with dermoscopy:** NR # **Dermoscopy:** 7FFM **Method of diagnosis:** Image-based **Prior test data:** Unclear whether clinical image also shown at same time **Diagnostic threshold:** >=2 **Test observers:** as described for Visual Inspection (above)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes

For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histology alone; no further details **Target condition (Final diagnoses)** Melanoma (invasive or *in situ*) 76 (8/468 melanomas in full sample were in situ; NR for <= 6 mm group) Benign nevi 524
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** None reported **Time interval to reference test**: NR
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	Unclear
Was the interval between application of the index tests less than one month?	Unclear
Are there any concerns that the test comparison could have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	
Are there concerns that the test comparison differs from the review question?	

Notes

Notes	

Benelli 2001

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** Not reported - only dates of training course and agreement study given (April-May 1999) **Country** Italy
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Slides of pigmented skin tumours were slected for evaluation during a training course on dermoscopy. Lesions not located on head, palms or soles histological slide available **Setting:** Training images; Authors institution. Institute of Dermatologic Sciences, University of Milan **Prior testing:** Slides of pigmented skin tumours were selected for evaluation during a training course on dermoscopy **Setting for prior testing:** Unspecified **Exclusion criteria:** None reported

	Sample size (patients): Not reported **Sample size (lesions):** No. included: 49 (paper reports 50 but only 49 accounted for in text) **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** ABCDE **Method of diagnosis:** Clinical photographs **Prior test data:** No further information used **Diagnostic threshold:** ABCDE Score >/=2; presence of 2 criteria; ABCDE Score >/=3; presence of 3 criteria. All criteria described in full **Diagnosis based on:** Single (n=1); Average (n=65; attending one of three courses in dermoscopy held to inform dermatologists about a new dermatoscopic diagnostic method (7FFM)) **Observer qualifications:** Dermatologists **Experience in practice:** Expert author; Not described for participating dermatologists **Experience with dermoscopy:** Expert author; Prior experience not described for participating dermatologists; all underwent dermoscopy training for study purposes # **Dermoscopy:** 7FFM **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used although clinicians had evaluated clinical images for the same 50 lesions earlier the same day **Diagnostic threshold:** Malignant if 7FFM Score >/=2; i.e. presence of one major feature or concurrent presence of two minor features. All criteria described in full **Test observers:** as described for Visual Inspection (above) # **Dermoscopy training:** 3 one day dermatoscopy courses held to inform dermatologists about authors' own new dermoscopy algorithm (7FFM). Each course lasted 6 hours. Morning session participants executed pre-test interpretation of clinical images using ABCDE. Then principles of dermatoscopy were presented during the course and as post-test, participants evaluated 50 dermoscopic slides of same lesions using 7FFM **Length of training** 1 day (6 hours) **Post-training experience:** <6months **Training format** In-person teaching

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 12/49 melanomas (paper reports 50 but only 49 accounted for in text) **Target condition (Final diagnoses)** Melanoma (invasive): 10; Melanoma (in situ): 2; BCC: 2 pigmented BCC 3 seborrhoeic keratoses, 2 pigmented basal cell carcinoma, 1 blue nevus, 2 angiokeratoma, 5 Spitz nevus, 5 junctional nevi, 9 compound nevi, 10 nevi undergoing regression.
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** None reported **Time interval to reference test**: Unclear
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	**Blinding between tests:** Clinical images interpreted in the morning and dermoscopic images in the afternoon **Time interval between index test(s):** image capture NR
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	Unclear
Was the interval between application of the index tests less than one month?	Unclear
Are there any concerns that the test comparison could have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Binder 1994

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** not reported **Country** Austria **Test set derived** From a sample of 200 PSL, two databases were randomly created for learning and testing purposes. The database was also provided with the histological diagnosis.
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Images of pigmented skin lesions randomly selected from a pigmented skin lesion image database. **Setting:** Secondary (general dermatology) **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** NR **Sample size (lesions):** NR **Participant characteristics:** NR **Lesion characteristics:** NR
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** (Modified) pattern analysis **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used; no additional clinical information was provided **Diagnostic threshold:** Observer correct diagnosis of melanoma; presence/absence of 8 ELM criteria were judged (pigment network, brown globules, radial streaming, pseudopods, black dots, margin regularity, pigmentation, depigmentation) and individual diagnosis made. **Diagnosis based on:** Consensus (2 observers); n=3. Images were examined independently by each observer; presence/absence of each ELM criterion decided by agreement of at least 2/3 observers **Observer qualifications:** Dermatologist **Experience in practice:** High experience or 'Expert' **Experience with dermoscopy:** High experience - described as 'ELM experienced dermatologists' **Any other detail** The images were obtained by photographing the PSL on 24x36 mm colour slide film, with oil immersion, using a Wild binocular stereomicroscope M 650 (Wild Heerbrugg AG, Switzerland) at a final magnification of x16 using flashlight illumination.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (no further details) Disease positive: 40; Disease negative: 60 **TARGET CONDITION (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 40 Benign naevus: 60
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	Excluded participants: **none reported** Time interval to reference test: **not reported** Time interval between index test(s): **not reported**
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Binder 1995

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR **Country** Austria
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions with available dermoscopy images, both with and without oil immersion, and histological confirmation of diagnosis. **Setting:** Secondary (general dermatology) **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** NR

	Sample size (lesions): No. included: 240 **Participant characteristics:** NR **Lesion characteristics:** Median thickness 0.7mm, IQR 0.48 to 0.76mm; all less than 1cm diameter

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images of lesions with and without oil immersion (*results *with* oil immersion used for primary analysis); images randomly presented to prevent consecutive presentation of slides for the same lesion. Each image was shown for 20 seconds with a 20 minute break after 240 slides **Prior test data:** No further information presented **Diagnostic threshold:** Correct diagnosis of melanoma. For each PSL image only one diagnosis was allowed (MM or not MM) **Diagnosis based on:** Average (n=19); 6 ELM experts and 13 randomly picked dermatologist 'nonexperts' **Observer qualifications:** Dermatologist **Experience in practice:** High - all certified dermatologists, experienced in clinical diagnosis **Experience with dermoscopy:** Mixed. 'Nonexperts' had no formal ELM training; 'Expert' users had been working scientifically in the development of ELM for at least 3 years **Any other detail** Images were obtained by photographing the PSLs on 24X36-mm color-slide film with ELM and without oil immersion (surface microscopy ISM]) using a binocular stereomicroscope (M 650, Wild AG, Hcerbrugg, Switzerland) at a final magnification of X 16 using flashlight illumination

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described) Disease positive: 57; Disease negative: 183 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 57; BCC: 8 Severe dysplasia: 42; other 'Benign' : 133
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Reference interval:** appears consecutive; "After photographing, all lesions were excised" **Excluded participants:** None reported
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Binder 1999

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** NR **Country:** Austria

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Randomly selected, histologically proven pigmented skin lesions with digital dermoscopy images **Setting:** Secondary (general dermatology) **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** NR **Sample size (lesions):** No. included: 250 **Participant characteristics:** NR **Lesion characteristics:** Thickness; 7 (17%) of the 41 melanomas were *in situ* lesions, 24 (59%) <0.75 mm, and 10 (24%) ranged from 0.76 to 1.8 mm.; Lesion size: all <=8mm diameter

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** ABCD; Pattern analysis/no algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used. Computer presented images in random order. **Diagnostic threshold:** ABCD classification (score> 5.45, >4.75); sensitivity and specificity also estimated at Q*. Subjective diagnosis (based on certainty of melanoma between 1 and 5) also recorded using pattern analysis (experts) or subjective rating (first-year residents) **Diagnosis based on:** Average (n=17) **Observer qualifications:** Dermatology residents - 5; Dermatologist (board-certified) - 12 **Experience in practice:** Mixed. First-year residents (n=5); practicing board-certified dermatologists with experience ranging from 4 to 15 years (n=8), and 4 board-certified recognized as experts mainly working at PSL units (n=4) **Experience with dermoscopy:** Mixed experience "Ten of the 17 raters (58.8%) reported on previous usage of the ABCD score, at least for testing purposes of the method." # **Dermoscopy training:** Written materials "Before testing all readers were instructed how to apply the ABCD criteria according to the literature published"

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 41 (16.4%) lesions; Disease negative: 209 (83.3%) lesions **Target condition (Final diagnoses)** Melanoma (invasive): 34 lesions; Melanoma (in situ): 7 lesions

	Benign naevus: 96 nevocellular nevi of the compound type, 62 junctional type, 24 dermal type, 13 Spitz nevi; 14 lentigines

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index to reference interval:** Consecutive; "After photography all lesions were excised" **Time interval between algorithms:** same time; image-based

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Blum 2003

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** Nov 1998 to March 2000; lesions overlap with Blum 2004b; data only included in algorithm comparison and not in primary analysis. **Country:** Germany **Test set derived:** Study develops a simplified version of ABCD algorithm; describes full data set "randomly divided into 2 groups (N0 and N1)" but new algorithm development was based on full dataset

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic skin lesions to be excised because of clinically and/or dermoscopically clear or suspicious malignancy, or by the wish of the patient after clear benign diagnosis **Setting:** Secondary (general dermatology) **Prior testing:** Clinical and/or dermatoscopic suspicion; Patient request for evaluation/excision **Setting for prior testing:** Not reported **Exclusion criteria:** Consecutive images of one lesion and external recorded images were not included. Images from all parts of the bodies were taken except of subungual and mucosal sites. **Sample size (patients):** 269 **Sample size (lesions):** 269 **Participant characteristics:** Male: (45/84) **Lesion characteristics:** Median Breslow thickness 0.96mm (SD 0.70mm) for all melanomas

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** modified ABCD (with and without 'E' for evolution); denoted by authors as ABC-point list; plus 7FFM; 7 point checklist; Menzies criteria; original ABCD not included due to lesion overlap with Blum 2004b **Method of diagnosis:** Dermoscopic images **Prior test data:** Unclear; study describes image acquisition and storage but does not described image interpretation **Diagnostic threshold:** Not reported for established algorithms "performed according to the criteria given in literature". For ABC-point list: >=4 points. A - Asymmetry of the outer shape in at least 1 axis (+1) (as per (Stolz; Nachbar); (A) - Asymmetry of the differential structures inside the lesion in at least 1 axis (+1) (new item); B - Abrupt cutoff of network at the border of the lesion in at least 1 quarter of the circumference (+1); C - Three or more colors (+1); D - Three or more differential structures (+1); E - Evolution/change noticed by the patient during the last 3 mo (+1); No or uncertain information +0; No change in the last 3 mo (-1) **Diagnosis based on:** Unclear (n=NR) **Observer qualifications:** Not reported; likely dermatologists **Experience in practice:** Not described **Experience with dermoscopy:** Not described.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	No
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Yes
Could the conduct or interpretation of the index test have introduced bias?	High risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone **Target condition (Final diagnoses)** Melanoma (invasive): 71; Melanoma (in situ): 9; Lentigo maligna 4 'Benign' diagnoses: 185
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** none reported **Time interval to reference test:** appears consecutive; consent given "for the recording and the following operation under local anaesthesia"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Unclear
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Blum 2003b

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** Sept 1998 - Dec 1999; lesions overlap with Blum 2004b; data only included in algorithm comparison and not in primary analysis. **Country** Germany
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** All lesions of patients with multiple atypical naevi excised due to suspicious clinical and/or dermoscopic features were included **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** lesions located on soles, palms, subungual and mucosal sites were excluded **Sample size (patients):** No. included: 205 **Sample size (lesions):** No. eligible: 254/ No. included: 254 **Participant characteristics:** Median age: 39.2 (1.6-86.4y); Male: 97 (47.3%)

	Lesion characteristics: NR

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy.**New algorithm (based on criteria of Hofmann-Wellenhof 2001) **Method of diagnosis:** Dermoscopic images **Prior test data:** Unclear; looks like blinded test interpretation **Diagnostic threshold:** lesions were classified into six different types according to morphological criteria of the new classification of atypical naevi (Clark naevi): reticular, globular and homogeneous or combinations of two of these types (Hofman Wellenhof 2001). If reticular, globular and homogeneous structures were found in one melanocytic lesion, this lesion was classified as a three-structure type **Diagnosis based on:** Consensus (2 observers); n=2 **Observer qualifications:** Not reported; likely dermatologists - "All images were viewed by two investigators" **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	No
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	High risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 75 MM; Disease negative: 179 **Target condition (Final diagnoses)** Melanoma (invasive): 63; Melanoma (in situ): 12 Benign naevus: Recurrent naevus 6; Splitz or Reed naevus 6; Congenital naevus 4; Blue naevus 3; Naevus without dysplasia 64; Dysplastic naevus 96
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** none reported; **Time interval between index and reference:** Assumed consecutive; "All patients gave written informed consent for the digital documentation and the following operation under local anaesthesia"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Blum 2004

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective **Period of data collection** September 1998 to March 1999; lesions overlap with Blum 2004b; data only included in algorithm comparison and not in primary analysis. **Country** Germany

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions excised due to suspicious clinical and / or dermoscopic features **Setting:** Pigmented lesion clinic **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** "Consecutive (repeat) images of one lesion were not included"; Malignant epithelial tumours (basal cell carcinoma, squamous cell carcinoma) were excluded. **Sample size (patients):** No. eligible: 157/ No. included: 157 **Sample size (lesions):** No. eligible: 162/ No. included: 157 **Participant characteristics:** Median age: 38.9 years (2 to 87 years); 45.2% male **Lesion characteristics** No change in the past 3 months was reported by 87 (55.4%) patients, followed by an observed change in 39 (24.8%) patients and no clear clinical history was given by 31 (19.7%) patients.Lesion site: Face/Ears: 9 (5.7%); Trunk: 102 (65%); Limbs: 38(24.2%); Acral 6(3.8%), mucosal sites 2(1.2%); Lesion thickness ≤1mm: 23 CMs (2 in-situ, 29 invasive) median Breslow thickness 0.86mm (standard deviation 0.54 mm; range 0.30-40 mm)

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** Pattern analysis **Method of diagnosis:** Dermoscopic images **Prior test data:** Images interpreted with and without clinical information (clinical history, age, sex of the patients and location of the tumour). **Diagnostic threshold:** Diagnosis of suspect CM made when the level of suspicion was 'roughly 50% or more'. " Clinical history was scored as positive "when any morphological change was recognized by the patient in the past 3 months. Morphological changes included change in size, colour or shape or any sign of ulceration or spontaneous bleeding. Possible dermoscopic classifications were benign nevi, atypical nevi, cutaneous melanoma and other benign epithelial tumours (e.g. seborrhoeic keratosis, angioma)" **Diagnosis based on:** Single observer (n=3) **Observer qualifications:** Not described; likely dermatologists **Experience in practice:** Not described **Experience with dermoscopy:** High/Moderate/Low "Three investigators ... with different experiences in dermoscopy: excellent (A), average (B) and beginner (C)."

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone **TARGET CONDITION (Final diagnoses)** Melanoma (invasive): 29; Melanoma (in situ): 2 Benign naevus: 53; 59 dysplastic naevi; 13 'epithelial benign tumours'

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability

Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** Consecutive images of one lesion were not included - assumed to be repeated images of same lesion; 162 images originally with 5 excluded to give a total study number of 157 lesion **Index to reference interval:** Assumed consecutive; ""All patients gave their written consent for the digital documentation and the following operation under local anaesthesia"

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Blum 2004b

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective; dermoscopic images assessed remotely from the patient **Period of data collection** 11 Nov 1998 - 2 Mar 2000 **Country** Germany **Test set derived** For validation of a new CAD procedure the complete collection (837 melanocytic lesions) was divided into two equal random subgroups n1 (training set) and n2(test set).

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic skin lesions imaged prospectively at the Pigmented Lesion Clinic of the Department of Dermatology, University of Tuebingen, Germany. **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Not reported **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** images from mucous membrane areas were excluded **Sample size (patients):** NR **Sample size (lesions):** No. eligible: 837/ No. included: 837 **Participant characteristics:** NR **Lesion characteristics:** Median breslow thickness for all melanomas 0.78mm (range 0.10-3.50)

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** 7FFM; 7 point checklist; ABCD; Menzies criteria **Method of diagnosis:** Dermoscopic images. **Prior test data:** Not clearly reported; results using new CAD algorithm were "compared with established dermoscopic classification rules applied to the same image material as the diagnostic computer algorithm." **Diagnostic threshold:** Not reported; original algorithms cited "established dermoscopic classification rules"; authored confirmed published standard thresholds of the mentioned algorithms were used. **Diagnosis based on:** Single observer; n= 1 **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Not described, assumed High; "lesions were prospectively classified as benign or malignant melanocytic lesions by the principal investigator (A.B.)"

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus follow up *Histology* Disease positive: 84; Disease negative: 185 *Clinical FU plus histology of suspicious lesions* - unexcised lesions were analysed independently by two of the investigators 2-3 times in 6 months on the basis of dermoscopic criteria. These lesions were classified as benign without any suspicion of malignancy by dermoscopic criteria, and follow-up records for at least 6 months showed no evidence of malignancy; n=568 **TARGET CONDITION (Final diagnoses)** Melanoma (invasive): 71; Melanoma (in situ): 9; Lentigo maligna 4 'Benign' diagnoses: 766
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** none reported **Time interval to reference test:** appears consecutive; "After obtaining informed written patient consent, 269 melanocytic skin lesions were excised under local anaesthesia and the diagnosis was established by histopathology"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Yes
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Unclear
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Bono 2002

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** June 1998-March 2000 **Country** Italy **Test set derived** A training set was separately derived using data obtained from 237 previously studies lesions (ref Farina 2000)
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Cutaneous pigmented lesions with clinical and/or dermatoscopic features that suggested a more or less important suspicion for CM **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** Location/site of lesion - Awkwardly situated lesions eg interdigital space, ears, nose or eyelids. Lesions on scalp excluded due to hair interference with reflectance - lesion size obvious large, thick melanomas **Sample size (patients):** No. included: 298 **Sample size (lesions):** No. included: 313 **Participant characteristics:** Mean age: 40y (10-86y); Male: 122; 41% **Lesion characteristics:** Lesion site: Head/Neck: 3%; Trunk: 61%; Limbs: 36%; Thickness ≤1mm: 70% (46/66); for 55 invasive MM: median thickness 0.64mm, range 0.17-3.24mm. Median diameter: 11mm (3-31mm)

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm (Training in the unit is based on ABCD but subjective experience of the clinician used for diagnosis) **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Diagnostic threshold:** Clinical diagnostic criteria based on subjective experience; emphasise lesion colour over dimensions. Diagnosis of suspect CM made when the level of suspicion was 'roughly 50% or more'.ABCD (asymmetry, border, colour, dimension) criteria have been the basis of training at the unit, but is not implemented in diagnosis; preferred emphasis on colour rather than dimensional character **Diagnosis based on:** Single observer; (n=1) **Observer qualifications:** Surgical oncologists **Experience in practice:** High experience or 'Expert'; over 5 years **Experience with dermoscopy:** Assumed high experience; over 5 years # **Dermoscopy** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** Presence of at least one of the following criterion: radial streaming, pseudopods, grey-blue veil, regression and erythema, whitish veil, black dots at the periphery (if network present), thick irregular network or milky-red background with red dots. **Test observers** as described for Visual Inspection (above) Dermatoscopy performed by a hand-held monocular microscope equipped with an achomatic lens permitting a magnification of x10 (Heine Delta 10).

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone **TARGET CONDITION (Final diagnoses)** Melanoma (invasive): 55; Melanoma (in situ): 11; BCC: 6 'Benign' diagnoses: 241;151 compound naevus, 24 junctional naevus, 12 dermal naevus, 12 lentigo simplex, 10 dysplastic naevus, 8 spindle-cell naevus, 8 sebhorrheic keratosis, 5 blue naevus, 3 spitz naevus, 8 other.
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: NR **Interval between index and reference**: NR
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes

Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	Same clinician undertook both diagnoses (in-person) **Time interval between index test(s):** Appears consecutive but not fully clear

Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Bono 2002b

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** Dec 2000 and Aug 2001 **Country** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Consecutive cutaneous pigmented lesions that were <=6mm in diameter and required surgical biopsy for diagnosis based on clinical or dermoscopic suspicion of CMM **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Not reported **Exclusion criteria:** lesion size >6mm; non-pigmented **Sample size (patients):** No. eligible: 349/ No. included: 157 **Sample size (lesions):** No. eligible: 375/ No. included: 161 **Participant characteristics:** Mean age 38y (14-82); Male: 61 (39%) **Lesion characteristics:** Site: head/Neck: 14 (9%); trunk: 88 (55%); limbs: 59 (36%); **Lesion size:** median: 5mm (1mm to 6mm)

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI).** No algorithm (ABCD (asymmetry, border, colour, dimension) criteria have been the basis of training at the unit, but is not implemented in diagnosis; preferred emphasis on colour rather than dimensional character) **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Diagnostic threshold:** A diagnosis of suspect CM is made when the level of suspicion is roughly 50% or more; lesions at a lower index of suspicion were considered benign for the purposes of this study. ABCD (asymmetry, border, colour, dimension) criteria have been the basis of training at the unit, but is not impleneted in diagnosis; preferred emphasis on colour rather than dimensional character **Diagnosis based on:** Single observer diagnostic criteria based on the subjective experience of the single clinician examining the pigmented lesion (n=2) **Observer qualifications:** surgical oncologists **Experience in practice:** High experience or 'Expert'; described as "expert in the recognition of pigmented lesions" **Experience with dermoscopy:** High experience /'Expert' users **Other detail:** Diagnostic criteria were based on the subjective experience of the single clinician examining the pigmented lesion, although the ABCD criteria have been the basis of training at the unit, they did not consider the ABCD mnemonic an essential formula for diagnosis of CM. They did not take into consideration the dimensional character and attributed great importance to the colour of a given lesion. # **Dermoscopy** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes in-person; dermoscopy performed by the same two clinicians who firstly made and registered the clinical diagnosis **Diagnostic threshold:** Dermatoscopic criteria for diagnosis of malignancy were radial streaming, pseudopods, grey-blue veil, regression and erythema, whiteish veil, black dots at periphery (if network present), thick irregular network, or milky-red background with red dots. A lesion was suspected for CM when positive for at least one criterion. **Test observers** as described for Visual Inspection (above) Any other detail This technique was performed by a hand-held monocilar microscope equipped with an achromatic lens permitting a magnification of 10x (heine Delta 10).

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear

For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 13 CM; Disease negative: 148 n **Target condition (Final diagnoses)** Melanoma (invasive): 10**;** Melanoma (in situ): 3; BCC: 2(1.2%) Mild/moderate dysplasia: 26 (16.1%); Sebhorrheic keratosis: 4(5%); Benign naevus: compound nevus 57(35.4%), junctional nevus 38(23.6%), spindle-cell nevus 6(3.7%), spitz nevus 5(3.1%), blue nevus 2(1.2%), other 6(3.7%), Lentigo simplex 2 (1.2%)
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	Excluded participants: none reported Time interval to reference test: not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	Dermoscopy performed by the same two clinicians who firstly made and registered the clinical diagnosis Time interval between index test(s): appears consecutive
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	Yes
Are there concerns that the test comparison differs from the review question?	Low concern

Notes

Notes	

Bono 2006

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective **Period of data collection:** Jan 2003 - Dec 2004 **Country:** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Consecutive patients with pigmented skin lesions with a maximum diameter of <=3mm undergoing excision. The decision for diagnostic excision was based on clinical and/or dermoscopic features suggesting a more or less important suspicion for CM **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) Melanoma and Sarcoma Unit; Istituto NazionaleTumori of Milan **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** - lesion size >3mm **Sample size (patients):** No. eligible: 204/ No. included: 204 **Sample size (lesions):** No. eligible: 206/ No. included: 206 **Participant characteristics:** Median age: 40 (6-74); Male: 71 (35%) **Lesion characteristics** Head/Neck: 8 (4%); Trunk: 84 (41%); Limbs: 114 (55%). Median size: 2mm (1 to 3mm)
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Diagnostic threshold:** A diagnosis of suspicious CM is made when the level of suspicion is roughly 50% or more; lesions at a lower index of suspicion were considered not CM; ABCD (asymmetry, border, colour, dimension) criteria have been the basis of training at the unit, but is not implemented in diagnosis; preferred emphasis on colour rather than dimensional character **Diagnosis based on:** Single observer; n= 1 **Observer qualifications:** Not reported (assumed Oncologist as per Bono 2002 and Bono 2002b); "single clinician examining the pigmented lesion" **Experience in practice:** Not described **Experience with dermoscopy:** Not described # **Dermoscopy** Menzies criteria **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** Dermoscopic criteria for diagnosis of malignancy were those of Menzies et al. (1996, 2003) **Test observers** as described for Visual Inspection (above) A hand-held monocular microscope equipped with an achromatic lens permitting a magnification of 10x (Heine Delta 20 microscope; Heine Ltd, Herrsching, Germany).

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: The slides were evaluated according to widely accepted criteria for the histopathological diagnosis of the various pigmented lesions. Disease positive: 23; Disease negative: 183

	Target condition (Final diagnoses) Melanoma (invasive): 19 (9.2%); Melanoma (in situ): 4 (0%) Mild/moderate dysplasia: dysplastic naevus 10 (4.9%); junctional naevus 76 (36.9%); compound naevus 50 (24.3%); dermal naevus 12 (5.8%); blue naevus 11 (5.3%); reed naevus 7 (3.4%); spitz naevus 3 (1.5%); halo naevus 3 (1.5%); lentigo simplex 7 (3.4%)Other 4 (1.9%)

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	No
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** none **Time interval to reference test**: not reported

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	Single observer performed both tests Time interval between index test(s): not reported

Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	Yes
Are there concerns that the test comparison differs from the review question?	Low concern

Notes

Notes	

Bourne 2012

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** June 1 - July 6 2009 **Country** Australia

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** All skin lesions consecutively excised at a skin cancer practice to exclude skin cancer and common lesions assessed as clearly benign and not biopsied were included **Setting:** Private; "a dedicated skin cancer practice in Brisbane, Australia" **Prior testing:** Clinical and/or dermatoscopic suspicion. Prior testing to assemble the test set occurs in secondary care by an experienced skin cancer doctor, then the images are tested on primary care professionals **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** clinically obvious basal cell carcinomas which could be easily diagnosed without dermoscopy were not included in the collection set. **Sample size (patients):** No. eligible: 46/ No. included: 46 **Sample size (lesions):** No. eligible: 50/No. included: 50 **Participant characteristics:** Mean age: 58 (30 -60); Male: 22 **Lesion characteristics:** Face = 8; Neck = 1; Chest = 3; Back = 21; Shoulder = 2; Arm = 3; Thigh = 4; Leg = 7; Foot plantar = 1

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** Clinical photographs **Prior test data:** No further information used; Image assessments were done on four occasions, each time using a different diagnostic approach. **Diagnostic threshold:** Not reported clinicians provided with Excel answer sheets for each method listing the various criteria used in that algorithm but no algorithm was cited for VI **Diagnosis based on:** Average (n=4) **Observer qualifications:** 3 GPs and 1 clinical nurse **Experience in practice:** Mixed; described as varying levels of dermatoscopic experience **Experience with dermoscopy:** Mixed; described as varying levels of dermatoscopic experience #

	Dermoscopy 3-point rule; Menzies criteria
	Method of diagnosis: Dermoscopic images
	Prior test data: No further information used; Image assessments were done on four occasions, each time using a different diagnostic approach.
	Diagnostic threshold: Not reported in paper; author communications states that standard thresholds were used - >=2 for the 3-Point checklist and Menzies method as described in original paper
	Test observers as described for Visual Inspection (above)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus other Histopathological examination (n=46); Expert diagnosis as benign (n=3); Digital follow up (n=1) **Target condition (Final diagnoses)** Melanoma (invasive): 1; Melanoma (in situ): 7; BCC: 6; Lentigo maligna 1 Sebhorrheic keratosis: 5; 'Benign' diagnoses: Banal nevus 10, Blue naevus 1, Nevus and seborrhoeic keratosis/solar lentigo collison 3, Solar lentigo 4, LPLK 4, Dermatofibroma 1, Psoriasis 1, Solar keratosis 2, Intraepidermal carcinoma 3, Regressed keratoacanthoma 1
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	No
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	High

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: As two of the methods (Menzies and 3 point checklist) related to only pigmented lesions, the 5 non-pigmented specimens in the set of 50 were excluded from the contingency tables for these methods. **Time interval to reference test**: "all skin lesions consecutively excised to exclude skin cancer were recorded"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Unclear
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	tbc
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	Yes
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Broganelli 2005

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** 1998-2002 **Country** Italy

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions excised at Dept of Dermatology; all lesions considered suspicious on clinical parameters (on at least one of ABCDE parameters apart from diameter) underwent dermoscopy; 2x2 for melanocytic only included **Setting:** Secondary (general dermatology) **Prior testing:** Clinical suspicion only; decision to excise "follows the dermoscopic diagnosis" **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** None reported **Sample size (patients):** NR **Sample size (lesions):** No. included: 638 melanocytic lesions **Participant characteristics:** Age range: between 2 months and 90 years **Lesion characteristics** NR

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy;** 7-point checklist **Method of diagnosis:** Unclear. Study describes 'day-to-day' office activity, but ELM interpretation referred to as evaluating 'recorded images' to split into melanocytic and non-melanocytic lesions. "Melanocytic lesions were investigated on the basis of a pattern analysis and those that revealed altered dermoscopic parameters were distinguished between minor and major criteria" **Prior test data:** Unclear what additional information was available **Diagnostic threshold:** > 1 alteration in minor criteria or >= 1 major char present; not further described. Based on data in Argenziano 1998, this is akin to a score of >=2 as major criteria score 2 points and minor ones score 1 each **Diagnosis based on:** Unclear appears to be in clinic diagnoses n=: NR **Observer qualifications:** Not reported likely dermatologists **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** - Histological diagnosis alone Details: "lesions were fixed with formaline and included in paraffin for histological examination. For some of them serial sections were made" **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 108 'Benign' diagnoses: non-melanomas= 530

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** none reported **Time interval to reference test:** not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Carli 1994

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Unclear **Period of data collection** November 1993 and May 1994 **Country** Italy
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Clinically suspicious melanocytic lesions undergoing excision for diagnostic purposes **Setting:** Secondary (general dermatology) **Prior testing:** Clinical suspicion of malignancy based on: recent lesion changes or presence of at least two of: diameter > 6 mm, asymmetric, irregular feathery edges, uneven or 'very' dark colour, 'increased or disappearance of skin outline'. **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** Clinically obvious melanomas excluded **Sample size (patients):** No. included: 67 **Sample size (lesions):** No. included: 67 **Participant characteristics:** Mean age 36y; median age 33; all >20 years; Male: 31% **Lesion characteristics:** None reported
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** Pattern analysis; criteria derived from a number of other studies (citations include Steiner 1993, Pehamberger 1987, Steiner 1981, Nachbar 1994, Bahmer 1990, Kenet 1993, Stolz 1989, Soyer 1987, Dal Pozzo 1993) **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination **Diagnostic threshold:** A pigment network that was irregular, accentuated, wide-meshed, with distinct borders, plus at least one of the following parameters :- Inhomogeneous depigmentation present at the periphery;- Presence of unevenly distributed black dots;- Uneven brown globules, with irregular distribution;- Presence of radial streaks;- Presence of pseudopods;- The presence of gray-blue veil. **Diagnosis based on:** Consensus (2 observers); n=2 **Observer qualifications:** NR; likely dermatologist **Experience in practice:** High - described as "two experienced observers" **Experience with dermoscopy:** High - as above

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	

Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** *Histology (not further described)* Disease positive: 5; Disease negative: 63 **Target condition (Final diagnoses)** Melanoma (invasive): 3; Melanoma (in situ): 2 'Benign' diagnoses: Atypical melanocytic hyperplasia 2; Nevi with architectural atypia 14; Nevi with 'cyto'-architectural atypia 7; No atypia 40
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Time interval to reference test:** ELM performed at the time of excision of the lesion
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Carli 2002

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective for clinical examination and in-vivo dermoscopy; retrospective image selection / Prospective interpretation for ex-vivo dermoscopic evaluation **Period of data collection:** June 1997 - December 1998 **Country:** Italy
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Clinically equivocal and suspicious pigmented skin lesions subjected to excisional biopsy at the Institute of Dermatology **Setting:** Secondary (not further specified) **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Secondary **Exclusion criteria:** None reported **Sample size (patients):** NR **Sample size (lesions):** 256 **Participant characteristics:** None reported **Lesion characteristics** Of the cutaneous melanomas, 14 (25.9%) were in situ melanoma (Clark level I), 18 (33.3%) were invasive with less than 0.75 mm thickness, 19 (35.3%) were of intermediate thickness (0.76–1.50 mm) and three (5.5%) were thicker than 1.5 mm. The median thickness of invasive melanomas was 0.94 mm ± 0.5 (SD) (range 0.2–6).
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** Unclear **Diagnostic threshold:** Not reported **Diagnosis based on:** Consensus (2 observers); final clinical diagnosis was based on agreement between the two observers. In case of disagreement, the opinion of a third observer was considered to be the judge for the diagnosis **Observer qualifications:** Dermatologist **Experience in practice:** High experience or 'Expert'; described as "dermatologists with extensive experience in both clinical and dermoscopic diagnosis of pigmented skin lesions" **Experience with dermoscopy:** High experience /'Expert' users # **Dermoscopy** Pattern analysis **Method of diagnosis:** In person diagnosis and image-based diagnosis. Clinical examination and in vivo dermoscopy were performed before excision by two trained dermatologists and diagnosis reached. Dermoscopic images were re-analysed by the same two observers at the end of the inclusion period (December 1998), blind to the previous clinical and histological diagnoses. **Prior test data:** N/A for in-person; For image-based: slides of dermoscopic images were evaluated using a viewer that made it impossible to analyse the clinical features of the lesion; both observers had access to clinical information, including the age of the patient, the site of the lesion, the history of change over time as reported by the patient at the time of in vivo examination. **Diagnostic threshold:** dermoscopic diagnosis was based on the ELM pattern analysis criteria, using the same diagnostic categories used for clinical diagnosis; characteristics investigated included pigment network, pigmentation, hypopigmentation, brown globules, black dots, pseudopods, radial streaming, grey-blue veil, atypical vascular pattern **Test observers** as described for Visual Inspection (above)

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone **Target condition (Final diagnoses)** Melanoma (invasive): 40; Melanoma (in situ): 14 BCC: 5 Sebhorrheic keratosis: 4; Benign naevus: 90 common melanocytic naevi; 78 melanocytic naevi; 9 blue naevi; 16 Spitz reed naevi
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	Excluded participants: none reported Time interval to reference test: not reported

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	In person clinical examination and dermoscopy Time interval between index test(s): the interval between the time in-vivo dermoscopy and re-evaluation of dermoscopic images was reported as 1 year

Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Carli 2002b

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Not reported **Period of data collection:** NR **Country:** Italy

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Clinically suspicious or equivocal pigmented skin lesions undergoing excision for diagnostic purposes; only lesions with a diameter of 14 mm or less were included **Setting:** Secondary (general dermatology) **Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** None reported **Sample size (patients):** No. included: NR **Sample size (lesions):** No. included: 57 **Participant characteristics:** None reported **Lesion characteristics:** thickness ≤1mm: 11 cases (5 in situ 6 invasive); All <=14mm diameter

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** Clinical photographs; Fixed focus distance of 10cm; images observed using a viewer in two separate diagnostic sessions **Prior test data:** No further information used; Contact (dermoscopic) images viewed first and then distant images (clinical), without knowing the classification of the contact image of the individual lesions. **Diagnostic threshold:** Not reported **Diagnosis based on:** Consensus (2 observers); n=2 **Observer qualifications:** Dermatologist **Experience in practice:** High experience or 'Expert'; States 'with experience in the field of PSL' **Experience with dermoscopy:** High experience /'Expert' users; 'experienced in the field of PSLs' **Other detail:** Any other detail Used an AF micro Nikkor 60 lens objective mounted on a Nikon f50 camera, with a fixed focus distance of 10cm # **Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used; Contact (dermoscopic) images viewed first and then distant images (clinical), without knowing the classification of the contact image of the individual lesions. **Diagnostic threshold:** Not reported **Test observers** as described for Visual Inspection (above) **Any other detail** Dermaphot device placed directly on the lesion without previous application of oil; only lesions with a diameter of 14 mm or less were included in the study. The image has an automatic, original magnification of x 10.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** *Histology (not further described)* Disease positive: 21; Disease negative: 36 **Target condition (Final diagnoses)** Melanoma (invasive): 6; Melanoma (in situ): 5; BCC: 10 'Benign' diagnoses: 36
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** No exclusions reported **Time interval to reference test:** Photographic procedures performed consecutively prior to surgery
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	Photographic procedures performed consecutively prior to surgery
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	Yes
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Carli 2003

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR **Country** Italy (from authors' institution)
Was a consecutive or random sample of patients enrolled?	Unclear

Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic lesions <14 mm in diameter, excised because they were clinically suspicious or equivocal **Setting:** Secondary (general dermatology) **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** Non melanocytic lesions **Sample size (patients):** NR **Sample size (lesions):** No. included: 200 **Participant characteristics:** None reported **Lesion characteristics:** All < 14 mm in diameter
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Pattern analysis; 7-point checklist; ABCD **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used; "Dermoscopic images were examined using a viewer" **Diagnostic threshold:** For ABCD: >5.45; For the seven-point check-list: >=3; Pattern analysis: threshold not described **Diagnosis based on:** Single observer (n=5); Average also presented **Observer qualifications:** Dermatology residents working out of the pigmented lesion clinic: 3 working predominantly in the inpatient units and in mycology laboratories, 1 working in dermato-allergology and 1 in the general outpatient units of the dermatology clinic. **Experience in practice:** Low experience or recently qualified **Experience with dermoscopy:** Low experience / novice users - considered as 'Trained'; All had undergone training in dermoscopy; one had previously taken part in a study on dermoscopy based both on pattern analysis and on the ABCD rule while the others had had no previous experience in practical dermoscopy during work in other fields of dermatology. # **Dermoscopy training:** Length of training 8 h formal lessons plus Interactive CD of Dermoscopy **Post-training experience:** 4h practice at pigmented lesion clinic **Training format** In-person teaching; CD-Rom tutorial; **Any other detail:** images taken at x10 magnification using a Dermaphot (Heine Optotechnik, Germany) mounted on a Nikon F50 camera.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	No
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** *Histology (not further described)* Disease positive: 44; Disease negative: 156 **Target condition (Final diagnoses)** Melanoma (invasive): 30**;** Melanoma (in situ): 14 Benign naevus: 156
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Reference interval** not reported

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Carli 2003b

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** 1999-2001 **Country:** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Clinically difficult to diagnose or equivocal melanocytic lesions randomly selected from image database; all melanomas less than 1mm thickness. **Setting:** Secondary (general dermatology) **Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** >/=1mm thick melanomas, non-melanocytic lesions, easy to diagnose, dermoscopically peculiar lesions (eg Blue nevi or Spitz nevi) **Sample size (patients):** NR **Sample size (lesions):** No. included: 200 **Participant characteristics:** None reported **Lesion characteristics** ≤1mm thickness: 64; median thickness 0.3mm, 25th-75th centile 0.00-0.58mm; Mean diameter 7.4 (SD79) mm; Median: 7mm (2-16mm) **Any other detail:** Same lesions appear to be reported in De Giorgi 2011 but with a different set of 8 observers (De Giorgi 2011 excluded from review on this basis)

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** Clinical photographs **Prior test data:** No further information used **Diagnostic threshold:** Not reported **Diagnosis based on:** Average; n= 8 **Observer qualifications:** Dermatology registrar; 2 final year residents. Dermatologist 6 **Experience in practice:** Mixed experience - 2 senior experts, 4 practicing dermatologists, 2 last year resident dermatologists. Both latter groups formally trained in dermoscopy. **Experience with dermoscopy:** Classified as 'high' due to expertise/training in dermoscopy use **Other detail:** Any other detail Clinical photos using Nikon F40 with macro lens at 15cm. # **Dermoscopy.** No algorithm (own choice) **Method of diagnosis:** Clinical photographs and dermoscopic images **Prior test data:** Unclear **Diagnostic threshold:** Not reported. All observers familiar with pattern analysis, ABCD and 7-point checklist, each was free to choose method of choice **Test observers** as described for Visual Inspection (above)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 64; Disease negative: 136 **Target condition (Final diagnoses)** Melanoma (invasive): 40; Melanoma (in situ): 24 Other: 136 melanocytic nevi
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** No exclusions reported **Time interval to reference test:** Interval not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Carrera 2016

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation. Each PLC provided up to 50 lesions with a 1:3 ratio of melanomas to nevi. Each contributor randomly selected either polarized or non-polarized images based on 1:1 randomisation. Following exclusions, lesions were randomised into 12 image sets containing 39 (n = 8) or 40 (n = 7) unique lesions and 5 non-unique lesion images (2 melanoma, 3 benign) that were repeated in all sets. **Period of data collection** NR **Country** Multicentre (images contributed from PLCs in Australia, Austria, Germany, Italy, Spain, Switzerland, and the United States)
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Images of melanocytic lesions including melanomas with an unequivocal histopathologic diagnosis, and histopathologically verified nevi or nevi demonstrating stability under sequential dermoscopic imaging over time.

	Setting: Specialist unit (skin cancer/pigmented lesions clinic) 12 PLCs **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** Acral, mucosal, or facial sites excluded; non-melanocytic appearance; Lesions with equivocal (final) diagnosis after review of the pathology report or sequential imaging; **Sample size (patients):** No. eligible: NR; No. included: NR **Sample size (lesions):** No. eligible: 580 lesion images were contributed; No. included: 477 (103 excluded on review by Memorial Sloan Kettering Cancer Center investigators) **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** 3-point rule; 7-point checklist; ABCD; Menzies criteria; Chaos and clues **Method of diagnosis:** Dermoscopic images **Prior test data:** Clinical image - Participants examined the close-up clinical image of each lesion before viewing the dermoscopic image; Image contributors also asked to provide information on anatomical location, patient age and sex, and imaging modality (polarized vs nonpolarized) but unclear whether this information was provided to observers or not. **Diagnostic threshold:** Observers asked to evaluate "a comprehensive list" of dermoscopic structures abstracted from various algorithms; overlapping criteria were merged into 1 criterion. Criteria were grouped into (1) global pattern, (2) pattern organization, (3) symmetry of contour, (4) symmetry of pattern, (5) architectural disorder, (6) abruptness of lesion border, (7) colours, and (8) melanocytic structures, including network and vascular structures. Algorithm performance was retrospectively assessed based on the following thresholds:7-point checklist - >=3; CASH - >=6; Menzies – NR; ABCD - >4.75; 3-point checklist – NR; Chaos and clues NR. For NR thresholds, author communications state "We assessed all of the algorithms. There isn't a threshold for these algorithms, there are just published rules of usage that determine the benign/malignant classification of the lesion" **Diagnosis based on:** Consensus (>=50%) - when 50% or more of the observers identified a dermoscopic feature for a given study lesion, the attribute was considered present; n= 130 (240 participants registered via the IDS website for the study; 103 completed all available images in their data sets and 130 evaluated >=20 lesions) **Observer qualifications:** GP 24; Dermatology registrar 25; Dermatologist 73; 1 medical student and 7 'other' **Experience in practice:** Mixed - mean 12 (SD 8.7) years of dermatology experience **Experience with dermoscopy:** Mixed - 122 (93.8%) reported being comfortable using dermoscopy,and 121 (93.1%) were regular users of dermoscopy **Dermoscopy training:** Algorithm tutorials were created and posted by dermoscopic experts through the International Dermoscopy Society (IDS) website; review of these was encouraged but not mandatory.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus follow up Histology: All melanomas (n=119) and a proportion of benign lesions (n not reported) *Clinical FU plus histology of suspicious lesions:* Sequential dermoscopic imaging over time'; not further detailed; Length of FU NR; nevi required to be either histopathologically verified or to have demonstrated stability under sequential dermoscopic imaging over time. **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 119 Benign naevus: melanocytic nevus: 358
Is the reference standards likely to correctly classify the target condition?	Unclear
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	

Flow and timing	**Excluded participants:** Poor quality index test image as exclusion criterion **Time interval to reference test:** NR **Time interval between index test(s):** in-person; sequential

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Unclear
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Coras 2003

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** 16 month period. Does not say the date **Country** Germany

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions undergoing excision due to diagnosis of melanoma or atypical nevus, to rule out melanoma or at the patient's request. Paper states "Each of the three participating dermatologists in private practive sent their digital images via email attachment including anonymized identification to the department of dermatology. (face to face diagnosis) **Setting:** Secondary (general dermatology) (teledermatoscopy diagnosis); Private care- Face to face diagnosis **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. eligible: 90; No. included: 45 **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**In person assessment** (for those comparing FtF vs histology) **Method of diagnosis:** Participating dermatologists with experience in dermatoscopy established a clinical diagnisis based on pattern analysis after personal consultation with the patient in their private practice clinics. **Prior test data:** Not reported **Diagnostic threshold:** Not reported **Diagnosis based on:** Single **Number of examiners**: 3 **Observer qualifications:** Dermatologist (experts with great experience in dermoscopy) **Experience in practice:** High **Experience with index test:** High **Teledermatology** **Acquisition and transmission of images:** Each of the participating dermatologists acquired digital images after face to face consultation, and send them via an email attachment with corresponding patient data and medical history. **Nature of images used:** Clinical photographs and dermoscopic images **Any additional patient information provided:** Clinical examination and/or case notes **Observer qualifications (remote diagnosis):** Physician experienced in dermatoscopy **Diagnosis based on:** Single observer **Method of diagnosis:** A physician evaluated the images and made a diagnosis based on the images and history of the patient **Other detail:** The participating dermatologists used the same technical equipment for the acquisition of digital images.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes

For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard:** Histology **Details:** The histological diagnosis of majority of cases was performed at the Department of Dermatology Regensburg - No. patients/lesions: 45 patients; Disease positive: 16; Disease negative: 29 **Target condition (Final diagnoses)** - Melanoma (invasive): 16; 'Benign' diagnoses: 29
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	1. Excluded participants: They reported that many images were of poor quality (10) and that only 45 biopsies were done 50 patients who did not have histology excluded 2. Time interval to reference test:Unclear 3. Time interval between index test(s): most likely days (email transmission of images for remote assessment).
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Cristofolini 1994

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** October 1990-June 1991 **Country:** Italy
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients with pigmented lesions presenting during a campaign for the early diagnosis of cutaneous melanoma at the Dermatology Department in Trento **Setting:** Secondary (general dermatology) **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** Lesions that were not taken into consideration included benign lesions, naevi of Unna and Miescher types and naevi that showed no inclusion criteria at the ABCDE clinical examination **Sample size (patients):** No. eligible: 700 people; No. included: not reported **Sample size (lesions):** No. eligible: 220; No. included: 220 **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** ABCDE **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Diagnostic threshold:** lesions showing at least two of the ABCDE criteria all of which were shown the same diagnostic importance, were considered positive. **Diagnosis based on:** Unclear; n=4 **Observer qualifications:** Dermatologist **Experience in practice:** High experience or 'Expert'; all trained in the recognition of pigmented lesions during a training course about the clinical diagnosis of naevi and melanomas; all working in a department where the early diagnosis of melanoma had been dealt with for over 10 years. **Experience with dermoscopy:** High experience /'Expert' users **Other detail:** ABCDE criteria are (asymmetry in shape, border irregular and notched, colour mottled-haphazard display, dimension >6mm, evolution changes in pigmentation) # **Dermoscopy** Pattern analysis **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical evaluation directly followed by dermoscopy **Diagnostic threshold:** lesion positive for at least one criterion: irregular and multicomponent pigmentary network pattern, peripheral dark network patches, sharp network margins, pseudopods(if network present), radial streaming (if network present), black dots at periphery (if network present), blue-grey areas (if network present) and whitish veil (milky way, if network present). **Observers as described above.**

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone **Target condition (Final diagnoses)** TARGET CONDITION (Final diagnoses) Melanoma (in situ and invasive, or not reported): 33 Mild//moderate dysplasia: 23 dysplastic naevi; Sebhorrheic keratosis: 4; Benign naevus: 158 common naevus Other: 2 thrombosed angiomas
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** No exclusions reported **Time interval to reference test:** Not described **Time interval between index tests:** clinical evaluation directly followed by dermoscopy

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	Clinical evaluation directly followed by dermoscopy
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	Unclear
Are there concerns that the test comparison differs from the review question?	Unclear

Notes

Notes	

Dal Pozzo 1999

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective; dermoscopic images assessed remotely from the patient **Period of data collection** Jan 1992 to Jun 1997 **Country** Italy **Test set derived:** "Training set" 218 pigmented lesions classified as: 45 melanomas (19 of which in situ), 38 epithelioid and/or spindle cell nevi; 45 melanocytc nevi; 45 mainly dermal melanocytic nevi."Test set"; 713 pigmented skin lesions-melanocytic in nature consecutively observed
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions observed clinically and dermoscopically at the Institute of Dermatology Sciences University of Milan; all excised. **Setting:** Secondary (general dermatology) **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** NR **Sample size (lesions):** No. eligible: test set - 713 PSLs; No. included: 713 **Participant characteristics:** Not reported **Lesion characteristics:** Not reported
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** 7FFM (own new algorithm) **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used **Diagnostic threshold:** The lesions where the sum of the features gave a score >=2 were diagnosed as being malignant. **Diagnosis based on:** Consensus (3 observers); n= 3 **Observer qualifications:** Not reported; appears to be the three co-authors; likely expert dermatologists **Experience in practice:** Not described **Experience with dermoscopy:** Not described **Any other detail:** Training set of 218 pigmented lesions used to develop new algorithm. All dermoscopic features recorded. Statistical significance of each feature assessed using Chi square test and Fischer's exact test. Final features chosen according to reproducibility by different observers and relationships with histopathological criteria predictive of malignancy. Final algorithm: To diagnose melanoma the presence of one major feature or the concurrent presence of two minor features is regarded as sufficient. "We attributed a score of 2 to the major features and a score 1 to the minor features: major features are regression erythema, radial streaming, gray-blue veil, irregularly distributed pseudopods; minor features are unhomogeneity, irregular pigment network, sharp margin."

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 168; Disease negative: 545 **Target condition (Final diagnoses)** Melanoma (invasive): 139; Melanoma (in situ): 29; BCC: 1 Sebhorrheic keratosis: 3; Benign naevus: Junctional melanocytic nevi-92; Mainly junctional compound melanocytic nevi-37; Compound melanocytic nevi-224; Congenital melanocytic nevi-20; Melanocytic nevi showing regression and inflammatory infiltrate-102; Combined melanocytic nevi-8; Epithelioid and/or spindle cell nevi-53;Lentigo simplex-3; Black reticulated solar lentigo-1; Melanoacanthoma-1
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** none reported **Time interval to reference test:** none reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

di Meo 2016

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** February to December 2014 **Country** Italy
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic skin lesions that underwent excision **Setting:** Secondary (general dermatology) **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** acral and mucosal lesions; dysplastic nevi excluded; Disagreement between evaluators on tumour histological classification - lesions that did not meet at least two consents were excluded; Poor quality index test image (considered under flow and timing) **Sample size (patients):** No. included: 125 **Sample size (lesions):** No. included: 125 **Participant characteristics:** Mean age: men 44.6y and women 50.0y; Male: 61; 58% **Lesion characteristics:** Thickness ≤1mm: all 32 melanomas
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** 3-point checklist; scored 3-point '4-point checklist' (authors own scoring); CASH algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used **Diagnostic threshold:** 3-point checklist - >= 2 criteria present; CASH score>7; 4-point checklist >2 **Diagnosis based on:** Unclear; lesions were "randomly assessed by two independent dermatologists" not clear if average or consensus; n= 2 **Observer qualifications:** Dermatologist **Experience in practice:** High **Experience with dermoscopy:** High - dermatologists with over 7 years of experience in dermoscopy **Any other detail:** The 3-point checklist criteria: asymmetry in colour and/or structures in one or two axes, pigmented network with thickened lines and irregular distribution, and any blue and/or white structure within the lesion. CASH algorithm has four criteria: colour, architectural disorder, symmetry and homo/heterogeneity. Scoring described in detail. 4-point checklist - doubled all three criteria of the 3-point checklist and chose the one conferring more sensitivity, specificity and accuracy (symmetry parameter doubled).

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone All lesions were excised and independently analysed by two dermatopathologists. The diagnosis of dysplastic nevus was based on the histopathological diagnostic criteria set by the World Health Organization Melanoma Programme.3 It was considered as a benign lesion Disease positive: 32; Disease negative: 93 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 32 Mild//moderate dysplasia: 50; Benign naevus: 43
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** Dysplastic nevi (n=50) excluded from 2x2; Poor quality index test images - exclusion criterion **Interval between index and reference standard:** Not clearly described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Dolianitis 2005

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** July 2001 to June 2002 **Country** Australia

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Dermoscopy training study using a CD with five test sets of images, each with 40 images of melanocytic skin lesions. Only good-quality macroscopic and dermoscopic images were included. **Setting:** Specialist unit; Victorian Melanoma Service, Department of Dermatology, University of Melbourne **Prior testing:** unclear **Setting for prior testing:** Not reported **Exclusion criteria:** Nonmelanocytic lesions; Poor quality index test image. Only good-quality macroscopic and dermoscopic images were included, where the whole lesion was visible, including the entire periphery (considered under flow/timing) **Sample size (patients):** NR **Sample size (lesions):** No. eligible: 40; No. included: 40 **Participant characteristics:** None reported **Lesion characteristics:** ≤1mm thickness: 14 invasive melanomas; median 0.50 mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI):** No algorithm **Method of diagnosis:** Clinical photographs alone **Prior test data:** No further information used **Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone. **Diagnostic threshold:** Not reported **Diagnosis based on:** Average; 61 participants (invited to participate in a study comparing dermoscopic algorithms; advertised at several medical meetings and on a Web site for primary care physicians). **Observer qualifications:** 10 dermatologists, 16 dermatology trainees, 35 GPs **Experience in practice:** Mixed. Participant (volunteers) "had a range of experience levels with assessment of skin lesions [outlined in detail in the paper] .. and a significant number were novices in dermoscopy". Paper reports 82% of participants responded that they assessed at least 2-4 PSL per week. **Experience in dermoscopy:** Mixed (as above); some educational material provided # **Dermoscopy:** Pattern analysis; 7-point checklist; ABCD; Menzies criteria **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used. Macroscopic image not shown. **Diagnostic threshold:** ABCD rule--lesions scoring > 4.75 (i.e. lesions 'of concern' were considered test positive along with those considered to be melanomas, scoring >5.45); thresholds not reported for the other algorithms (original studies referenced). Test observers as described for Visual Inspection (above) # **Dermoscopy training:** Participants were given explanatory written material as well as 3 compact discs (CDs). Two CDs contained educational material on dermoscopy, one from the American Academy of Dermatology and the other from the Web site www.dermoscopy.org. Participants were advised to work through all the educational material prior to assessing the test set of images. **Length of training:** not clear **Post-training experience:** <6 months **Training format** Online/ Written materials/CD-Rom tutorial

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus other (one lesion described as having no biopsy performed) *Histology (not further described)* Disease positive: 20; Disease negative: 19 *Expert dx:* 1 **Target condition (Final diagnoses)** Melanoma (invasive): 18; Lentigo maligna 2 Benign naevus: 7 dysplactic nevi; 3 spitz nevi; 3 junctional nevi; 2 compound nevi; 4 other (ink-spot lentigo, blue nevus, solar lentigo, ephelis)
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	No
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	High

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: none reported **Time interval to reference test**: not reported **Time interval between index test(s)**: not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Unclear
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	Yes
Was the interval between application of the index tests less than one month?	Unclear
Are there any concerns that the test comparison could have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Dreiseitl 2009

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** Test set: Feb-Nov 2004 **Country** Austria **Test set derived** Study focuses on test set but gives detail of separate study in which classifier was trained
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients presenting at pigmented skin lesion clinic **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) The pigmented skin lesion unit of the Department of Dermatology at the Medical University of Vienna serves as a secondary and tertiary referral center. **Prior testing:** Not reported **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** None reported **Sample size (patients):** No. eligible: 511; No. included: 458 with complete information **Sample size (lesions):** No. eligible: 3827; No. included: 3021; however data reported on a per patient basis **Participant characteristics:** None reported **Lesion characteristics:** None reported
Are the included patients and chosen study setting appropriate?	Yes

Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	Low concern

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** In person diagnosis; physicians were instructed to perform an independent routine examination on the study participants **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** Not reported; decision to excise to rule out melanoma histopathologically **Diagnosis based on:** Single observer (n=1) **Observer qualifications:** Dermatologist; data reported for 6 additional less experienced observers using MoleMax II system (reported in CAD review) **Experience in practice:** High experience; "Expert dermatologist" **Experience with dermoscopy:** High experience

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus follow up *Histology (excision);* No. patient/lesions: Not reported *Clinical FU plus histology of suspicious lesions* Length of FU: 6 months; No. patients: Not reported **TARGET CONDITION (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 27 patients; 31 lesions 'Benign' diagnoses: 431 patients; 2990 lesions
Is the reference standards likely to correctly classify the target condition?	Unclear
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** 806 lesions (53 patients) with inadequate follow-up **Index test to reference standard interval:**
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Yes
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Duff 2001

Patient Selection

A. Risk of Bias	

Patient Sampling	**Study design:** Case series **Data collection:** Retrospective **Period of data collection** January 1993 to December 1998 **Country** UK

~~Was a consecutive or random~~ sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Excised lesions recorded on pigmented lesion clinic database with data supplemented with hospital PAS and pathology database. **Setting:** rapid-access PLC at Frenchay Hospital **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** None reported **Sample size (patients):** No. eligible: 9968 attended clinic during time period; No. included: NR **Sample size (lesions):** No. included: 2372 (1256 undertaken immediately) **Participant characteristics:** Male: 40% (n=950) **Lesion characteristics:** Mean thickness of melanomas reported graphically pa (all estimates are appoximate):1993 - 1.44mm; 1994 - 0.82mm; 1995 - 1.22mm; 1996 - 1.40mm; 1997 - 1.35; 1998 - 0.90mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy;** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** Not reported; Diagnosis of melanoma **Diagnosis based on:** Single observer; (n= 2 as reported in Kirkpatrick 1995) **Observer qualifications:** Plastic surgeons **Experience in practice:** NR **Experience with dermoscopy:** Not described; "A consultant examines all lesions with a dermatoscope."

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** *Histology alone;* histopathologist with special interest in melanoproliferative lesions Disease positive: 586; Disease negative: 1786 **Target condition (Final diagnoses)** Melanoma (invasive): 400; Melanoma (in situ): 186 (128 In situ 58 LMs) BCC: 316; cSCC: 97 Atypical/dysplastic 195; "other" 14; 'Benign': 1164
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias

Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not all lesions were excised immediately (2372 excisions were undertaken, of which 1256 were done immediately

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Dummer 1993

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective; dermoscopic images assessed remotely from the patient **Period of data collection** 12 month period (year/dates NR) **Country** Germany

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients with skin lesions difficult to diagnose clinically **Setting:** Secondary **Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) a type of specialist care- dermatology based clinic **Exclusion criteria:** Patients who had excisions performed in individual practices or where there was no histology or cases that were so obvious they didn't need to have further investigation (clearly benign) **Sample size (patients):** **Sample size (lesions):** No. eligible: 824; No. included: 771 **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection:** No algorithm **Method of diagnosis:** In person **Prior test data:** In person **Other test data:** Dermoscopic images viewed separately **Diagnostic threshold:** NR **Diagnosis based on:** Single observer; (n=2 or 3) **Observer qualifications:** Unclear; clinician based in Dermatology clinic (assumed Dermatologist) **Experience in practice:** Unclear **Experience with index test:** Unclear # **Dermoscopy:** Pattern analysis **Method of diagnosis:** Dermoscopic images **Prior test data:** Unclear **Diagnostic threshold:** Not reported **Observers:** as described above

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 23MM; Disease negative: 748 benign **Target condition (Final diagnoses)** Invasive melanoma: 23 Benign naevus 706; Sebhorrheic keratosis 4; Benign non-melanocytic naevus 32
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	No
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: 53 non-melanocytic lesions not included in the final analysis (no melanomas present in this group) **Time interval to reference test**: Not reported **Time interval between index test(s)**: Not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Durdu 2011

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** Jan 2006 to January 2009 **Country:** Turkey
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions that could not be diagnosed with only dermatologic physical examination **Setting:** Secondary (general dermatology) **Prior testing:** Clinical examination and dermoscopy **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** None reported **Sample size (patients):** No. included: 176 **Sample size (lesions):** No. included: 200 **Participant characteristics:** Mean age: 48y (4 to 85y). Male: 64; 36.4% **Lesion characteristics:** 9% nodulo-ulcerative, 56% papular, 17% macular, 10% nodular, 8% plaque.
Are the included patients and chosen study setting appropriate?	Yes

Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** ABCD **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination **Diagnostic threshold:** Two step process: step 1 melanocytic and non melanocytic were differentiated (Braun 2005; Zalaudek 2008); step 2 ABCD applied to melanocytic lesions only (threshold >5.45) **Diagnosis based on:** Single observer; n= 2; one for dermoscopy diagnosis and one for Tzanck smear **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (Excisional biopsies (n=166) or punch biopsy (n=34) Details: "Biopsy specimens were stained with hematoxylin and eosin. Immunohistochemical (anti-S-100 and human melanoma black [HMB]-45) and histochemical (Fontana-Masson) stains were also applied, if necessary"; interpretation by a 'pathologist' Disease positive: 46; Disease negative: 154 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 10; BCC: 34; 1 pigmented mammary Paget disease; 1 pigmented metastatic mammary carcinoma Sebhorrheic keratosis: 24; Benign melanocytic naevus: 100; Dermatofibroma 12; Warts 16; 1 Dirt; 1 hereditary hemorrhagic telangiectasia
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Time interval to reference test:** appears consecutive. Following dermoscopic examination and cytology "either a punch or an excisional biopsy specimen was taken from the lesions and was examined histopathologically"
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Feci 2015

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Randomised controlled trial of the effect of ambient stressors and time constraints on decision making; PSL images were randomised to control group, ambient stress group and time stress* group (*result included in main analysis) **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** Jan-Dec 2013 **Country:** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions suspicious for melanoma and with histopathological diagnoses **Setting:** Secondary (general dermatology) **Prior testing:** Clinical and/or dermatoscopic suspicion of melanoma or atypical **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** Not clearly reported however only melanomas and atypical nevi included **Sample size (patients):** No. included: Appears to be one lesion per patient - 'consecutive PSL removed from different patients' **Sample size (lesions):** No. included: 321; 102 in time stress group **Participant characteristics:** None reported **Lesion characteristics:** mean thickness 0.28 mm, range - in situ to 1.88 mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Pattern analysis **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used; "dermatologists "knew neither the aim of the study nor the number of nevi and melanomas within each sample group **Diagnostic threshold:** Not reported **Diagnosis based on:** Unclear; Appears to be single (and different) observer per arm of the trial (n=3). The time stress group "simulated clinical decision making by arbitrarily allowing a time of 10s for the evaluation of each PSL" using Microsoft Power Point slide show **Observer qualifications:** Dermatologist **Experience in practice:** High experience **Experience with dermoscopy:** High experience; described as 'expert dermatologists' 'with at least 10 years experience in dermoscopy' **Any other detail** Dermoscopic image acquisition was performed using DermLite ® II pro (3Gen; DermLite, San Juan Capistrano,Calif., USA) connected to a Cyber-shot 7.2 megapixel camera(Sony Inc., Tokyo, Japan).

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: Diagnosis was based on AJCC guidelines (Balch 2001) and always made by the same pathologist Disease positive: 102 (34 per arm); Disease negative: 219 (73 per arm) **Target condition (Final diagnoses)** Melanoma (invasive): 69 (33 per arm); Melanoma (in situ): 33 (11 per arm) Benign naevus: Benign melanocytic nevi 219

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes

Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** Appear to have excluded on image quality "Among 686 PSL dermoscopic images acquired during the study period, 321 were suitable for our study" **Time interval to reference test:** NR

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Feldmann 1998

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** Not reported **Country:** Austria

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic lesions examined by dermatoscopy prior to excision **Setting:** Secondary (general dermatology) **Prior testing:** Not reported; "selection for excision was not exclusively based on the dermatoscopic findings but also according to the wishes of the patients." **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. included: 500 **Participant characteristics:** None reported **Lesion characteristics:** Mean Breslow thickness 0.49mm, range 0.12 to 1.38mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** ABCD **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination **Diagnostic threshold:** From Nachbar's study >5.45; from study results >4.2 **Diagnosis based on:** Unclear; n= unclear **Observer qualifications:** Not reported **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Deatils: histology was performed with at least three incisions (naevi), and serial sections through the entire lesion (melanomas) . The assessment was based on the generally accepted criteria for dysplasia and malignancy [1, 4]. In the case of diagnostic uncertainties, the Austrian reference center for histopathological diagnostics carried out a second assessment. Disease positive: 30 MM; Disease negative: 470 **Target condition (Final diagnoses)** Melanoma (invasive): 25; Melanoma (in situ): 5 Mild//moderate dysplasia: 190; Benign naevus: 272; 7 Lentigines 1 lentigo nevi
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** Results not presented for 8 lesions **Time interval to reference test:** Appears consecutive; dermoscopy described as used "prior to ... excision and histolog(y)"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Unclear
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Ferrari 2015

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** 2010 **Country** Italy
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic lesions with equivocal clinical and/or dermoscopic features that underwent excision and had a complete set of dermoscopy and RCM images with histopathology report. [Only dermoscopically featureless (scoring 0-2 on 7-point checklist) or equicocal lesions (those scoring 3-4 on dermoscopy 7-point checklist) were included in RCM evaluation.] **Setting:** Secondary (general dermatology) **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** Incomplete histopathology report; 90 ‘positive-clear cut’ lesions (scoring 5 or more on 7-point checklist) were excluded from RCM evaluation Poor quality index test image "Only lesions with high quality dermoscopic images, a complete set of confocal images and histopathology report available were included in the study"; considered under flow and timing. **Sample size (patients):** No. included: NR **Sample size (lesions):** No. eligible: 322; No. included: 322 for dermoscopy; 232 for RCM **Participant characteristics:** None reported **Lesion characteristics:** Overall mean thickness 1.05 +/- 16 mm, range 0–10 mm (70 melanomas); Those scoring 0-2 on 7-point checklist: mean 0.18 +/- 0.42 mm; range 0–0.94 mm) (6 melanomas). Those scoring 3-4 on 7-point checklist: mean 0.36 +/-0.42, range 0-1.4mm (17 melanomas)
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** 7-point checklist **Method of diagnosis:** Dermoscopic images **Prior test data:** RCM and dermoscopy images interpreted by same observer; no indication of randomisation or interpretation in isolation **Diagnostic threshold:** 'Featureless' lesions for score ranging between 0 and 2, 'positive-borderline' lesions for score between 3 and 4 and 'positive- clear cut' lesions for score from 5 to 10. **Diagnosis based on:** Single observer (n= 1) **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Assumed to be High - Described as "dermatologist trained in dermoscopy and RCM"

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Deatils: Histopathology was performed by a Board Certified Pathologist Disease positive: 70; Disease negative: 252 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 70 'Benign' nevi: 252 (including 15 Spitz nevi)
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** "Only lesions with high quality dermoscopic images, a complete set of confocal images and histopathology report available were included in the study" **Time interval to reference test:** Images taken 'before excision' "Before excision, all lesions were recorded by means of digital dermoscopy and RCM"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Ferris 2015

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Unclear. Some dermoscopic images were collected prospectively and some were obtained from collection of existing images; selection process not described. **Data collection:** Retrospective image selection / Prospective interpretation

	Period of data collection not reported **Country** USA **Test set derived.** Study developed a new CAD classifier using training/test set of images; plus a Reader study* conducted to compare accuracy with dermatologist interpretation of images (*reported here). Some dermoscopic images used to train the classifier were obtained from publicly available or purchased image libraries, these were not included in the reader study or used to test the performance of the classifier. The image set was randomly divided into 2 by diagnosis, with half used for training and half used for testing, with the exception that all high-grade dysplastic nevi were exclusively assigned to the training set to increase the representation of dermoscopic features that could be present in melanoma. Results are extracted only for the test set.

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Unclear
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Dermoscopic images of skin lesions excised on the basis of clinical suspicion of malignancy, with available histologic diagnoses. Reader study included one melanoma that was misclassified as benign by the new CAD classifier plus random sample of images determined to be of suitable quality for display on a computer screen. **Setting:** Secondary (general dermatology) **Prior testing:** Clinical suspicion (no further detail) **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** High-grade dysplastic nevi were not included in the test set or Reader study **Sample size (patients):** No. eligible: not reported; No. included: not reported **Sample size (lesions):** No. eligible: 473 (includes 273 randomised to training set and 27 non-biopsied lesions); No. included: CAD test set 173 lesions; Dermscopy- 65 lesions **Participant characteristics:** None reported **Lesion characteristics:** Test set: mean lesion thickness 0.76 mm, median 0.5 mm, range 0.2-98 mm); Reader study: mean 0.93 mm, median 0.74 mm, range 0.2 to 98 mm.

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used **Diagnostic threshold:** Not reported **Diagnosis based on:** Average (n=30); 35 invited to participate. **Observer qualifications:** 2 board certified dermatologists, 10 dermatology residents, and 8 physician assistants currently practicing dermatology **Experience in practice:** Mixed **Experience with dermoscopy:** Mixed; all observers self reported some training and experience with the use of dermoscopy. Among board certified dermatologists, 67% reported using dermoscopy ‘‘always/almost always’’ or ‘‘very frequently.'', compared to 90% of the dermatology residents and 75% of the physician assistants.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: All lesions were biopsied based on clinical suspicion of malignancy. All histologic diagnoses were rendered by at least 1 board-certified dermatopathologist and were used as the reference standard for diagnosis Disease positive: Derm 25MM; CAD 39MM / Disease negative: Derm=40; CAD= 134 **Target condition (Final diagnoses)** For Reader study only: Invasive melanomas 15; Melanoma in situ 10 in situ Low-grade dysplastic nevi 16, benign nevi 14 , blue nevi 2, lentigines 4 , seborrheic keratoses 4
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** none reported **Time interval to reference test:** 'Dermoscopic images of skin lesions were collected before biopsy'
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Friedman 2008

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR; lesions selected in July 2005 **Country** US **Test set derived** MelaFind data randomly split into training and test sets however Melafind has previously been evaluated, the only difference here being that only small lesions were included. Full dataset included in review
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** A database of images of pigmented skin lesions <=6mm was used to sample images of melanoma and non melanoma lesions; "approximately 80% of the lesions were biopsied to rule out melanoma, whereas the remaining lesions were biopsied mostly to rule out nonmelanoma skin cancer or because of patient concern." **Setting:** Mixed (private and secondary); digital dermoscopic database acquired by Electro-Optical Sciences Inc for the development and testing of MelaFind; 26 clinical sites have contributed (Dermatologic hospital-based clinics and private practice offices). **Prior testing:** Selected for excision (no further detail). All lesions excised or underwent shave biopsy **Setting for prior testing:** Not reported **Exclusion criteria:** High-grade dysplastic nevi were excluded. Previously biopsied, ulcerated, or bleeding lesions also excluded, as were those on mucosal surfaces and lesions that contained foreign matter (eg, tattoos). **Sample size (patients):** No. included: 94 **Sample size (lesions):** No. eligible: 1977; No. included: 99 **Participant characteristics:** None reported **Lesion characteristics** 21 invasive MM: median thickness 0.32mm (0.10, 1.40mm). Lesion size: Range: 2mm to 22mm
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images. Readers were provided with a CD-ROM with colour dermoscopic images created using MelaFind multispectral image; for some cases standard dermoscopic images were also available. The equivalence of the two image types was assessed for a sample of 10 lesions by 3 readers. **Prior test data:** Readers provided with participant gender, age, and lesion location; All evaluations were performed independently. **Diagnostic threshold:** Clinical diagnosis; “Is this lesion a melanoma?” and “Would you biopsy/excise this lesion?”. If readers indicated that they would biopsy the lesion because they were sure it was melanoma or to rule out melanoma, then the case was considered true positive (TP) **Diagnosis based on:** Average; mean and median reported (n=10); used mean value for review purposes **Observer qualifications:** 9 dermatologists; 1 nurse practitioner specializing in dermatology **Experience in practice:** High experience or ‘Expert’; "All 10 readers were expert dermoscopists (9 dermatologists and 1 nurse practitioner specializing in dermatology) **Experience with dermoscopy:** High experience /‘Expert’ users

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard:** Histological diagnosis alone Deatils: The original histology slides were evaluated by 2 out of 4 study dermatopathologists without knowledge of any additional clinical information; in cases of significant discordance in diagnoses, the slide was reviewed by a third study dermatopathologist. A lesion with at least 1 diagnosis of melanoma by the study dermatopathologists is considered melanoma. Dysplastic nevi with severe cytologic atypia were considered high grade, and those with mild to moderate atypia were considered low grade. Disease positive: 49; Disease negative: 50 **Target condition (Final diagnoses)** Melanoma (invasive): 21; Melanoma (in situ): 28; BCC: 2 Mild/moderate dysplasia: 32 low grade dysplastic; Sebhorrheic keratosis: 2; 14 other benign
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Timing between image acquisition and original histology not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Gereli 2010

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** not reported **Country:** Turkey
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Images of melanoma and nonmelanoma pigmented skin lesions; nonmelanoma lesions clinically considered to be atypical before dermoscopic examination and excisional biopsy. Atypicality was determined by the presence of at least three of the following features: a diameter greater than 5 mm, ill-defined borders, irregular margins, and the presence of papular and macular components. Melanoma and nonmelanoma lesions separately sampled **Setting:** Secondary (general dermatology) Auth inst: dept Dermatology, Istanbul, Turkey **Prior testing:** Clinical suspicion of malignancy **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** Not reported

	Sample size (lesions): No. included: 96 **Participant characteristics:** None reported **Lesion characteristics:** All > 5mm diameter

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** 3-point rule; 7-point checklist **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used **Diagnostic threshold:** 3-point rule: >=2 chars present (asymmetry, atypical pigment network, blue-white structures); 7-point checklist: >=3 chars present (Atypical pigment network, blue-whitish veil, atypical vascular pattern, irregular streaks, irregular dots/globules, irregular pigmentation, regression structures) **Diagnosis based on:** Average (n=3) **Observer qualifications:** Not reported; likely dermatologists (co-authors based in dept Dermatology) **Experience in practice:** Mixed: "two experienced and one inexperienced observers" **Experience with dermoscopy:** Mixed

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (no further details) Disease positive: 48; Disease negative: 48 **Target condition (Final diagnoses)** Melanoma (invasive): 44 (14 superficial spreading, 12 nodular, 10 acral, 4 lentiginous, 4 without classification of tumour thickness); Melanoma (in situ): 4 Sebhorrheic keratosis: 2; Blue nevi 2; Melanocytic nevi 44
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Gilmore 2010

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** 2003-2008 **Country** Austria **Test set derived:** Not reported. Training set: 65 melanomas and 65 dysplastic naevi, Test set:36 melanomas and 33 dysplastic naevi (included in review).

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Atypical melanocytic lesions with polarised dermoscopic images; describes database as a "random, but representative, cohort" but does not describe method of selection **Setting:** Secondary (general dermatology) **Prior testing:** Unclear **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** No. included: NR **Sample size (lesions):** No. included: 199: Derivation set n=130 Test set n= 69 **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used; described as blinded assessment **Diagnostic threshold:** Not reported - subjective impression; excise or not **Diagnosis based on:** Single observer (n=1) **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Not described; implies High or expert assessment. Conducted by one of the co-authors.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: "lesions were excised and examined microscopically by expert dermatopathologists using standard histopathologic diagnostic criteria" Disease positive: 36=test set and 65=derivation set Disease negative: 33=test set and 65=derivation set **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 36 test set and 65 derivation set Dysplastic naevi: 33 test set and 65 derivation set

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Glud 2009

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective; dermoscopic images assessed remotely from the patient **Period of data collection** Jan to Apr 2007 **Country** Denmark
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients referred for excision biopsy of pigmented lesions where the diagnosis of melanoma could not be excluded on clinical investigation **Setting:** Secondary (other); Dept Plastic Surgery and Burn Unit **Prior testing:** Clinical suspicion of malignancy **Setting for prior testing:** Secondary (not further specified) **Exclusion criteria:** None reported **Sample size (patients):** No. included: 65 **Sample size (lesions):** No. included: 83 **Participant characteristics:** Median age 47 yrs (18 to 90y); Male - 29; 45% **Lesion characteristics:** melanoma thickness 0.29 mm to 18mm
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used **Diagnostic threshold:** Not reported - diagnosis of melanoma **Diagnosis based on:** Single observer (n=1) **Observer qualifications:** Dermatologist **Experience in practice:** High **Experience with dermoscopy:** High experience; "dermoscopic images were examined by an experienced dermatologist" Any other detail The dermoscopic and SIAgraphic images were obtained by SIAscope II (Amon Clinica, Cambridge, UK) and stored using the proprietary Dermetrics software (Astron Clinica).

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: Following image acquisition "the excision biopsy was performed and an experienced histopathologist examined the tissue". Breslow thickness and Clark level were determined by standard histopathologic examination. Tumor staging was performed as described by Balch et al according to the 2001 melanoma staging system (Balch 2001). Disease positive: 12; Disease negative: 71 **Target condition (Final diagnoses)** Melanoma (invasive): 7; Melanoma (in situ): 5; 1 melanoma metastasis (incl as benign) Sebhorrheic keratosis: 1; Benign naevus: 57; 'Benign' diagnoses: bowens 1 haemangioma 1 lentigo simplex 2 epidermal naevi 2 DF 6
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Following image acquisition "the excision biopsy was performed"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Gokdemir 2011

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Not reported. **Period of data collection:** 2005-2009 **Country:** Turkey
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients with melanocytic and non-melanocytic skin lesions excised due to dermoscopic suspicion of malignancy or dysplasia . **Setting:** Secondary (general dermatology) **Prior testing:** Not reported **Setting for prior testing:** Unspecified **Exclusion criteria:** None reported **Sample size (patients):** No. eligible: 1264; No. included: 362 **Sample size (lesions):** No. included: 449 **Participant characteristics:** Mean age 40.3 yrs (+/- 1.08), range 1 to 89 yrs; Male: 160; 44.2% **Lesion characteristics:** None reported
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Unclear; appears to be in-person diagnosis

	Prior test data: Clinical examination **Diagnostic threshold:** Not reported; diagnosis of melanoma **Diagnosis based on:** Unclear (n=NR) **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** High experience - at least 2 years experience with Molemax II.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone; *not further described* Disease positive 13; Disease negative 433 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 13; BCC: 45 Benign: Not described
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Grimaldi 2009

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** Oct 2005 - Mar 2006 **Country** Italy
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Cutaneous pigmented lesions with digital images forwarded by primary care physicians to a referral centre for confirmation of diagnosis. **Setting:** Primary; Lesions selected for referral by GPs; accuracy of GP diagnosis assessed **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** Lesions whose removal had been explicitly demanded by the patients for aesthetic reasons, as well as those irritated or subjected to trauma **Sample size (patients):** No. included: 197 **Sample size (lesions):** No. included: 235 **Participant characteristics:** None reported **Lesion characteristics:** None reported
Are the included patients and chosen study setting appropriate?	Yes
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI);** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Other test data:** "two-step judgment (before and after dermoscopy) formulated by the sending physician, who labelled each lesion as 'benign' or 'suspicious for malignancy'." **Diagnostic threshold:** Not reported "Each physician was asked to formulate a written first judgment of every lesion before digital acquisition and to re-evaluate it after dermoscopy" **Diagnosis based on:** Single observer; (n= 13) **Observer qualifications:** GP; From approximately 250 primary care clinicians attending a conference, 13 volunteered to participate **Experience in practice:** Not clearly described; assumed to be Low experience with pigmented lesions **Experience in dermoscopy:** Unclear; classified as 'trained' - "simple protocols for diagnosis were made up and given to the participants via e-learning courses, direct meetings, and involving self assessment procedures" # **Dermoscopy** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** NR - "The evaluation method followed the ABCD rule of dermoscopy according to Nachbar et al"; not fully clear whether this relates to GP in-person diagnosis or telediagnosis at reference centre - "two-step judgment (before and after dermoscopy) formulated by the sending physician, who labelled each lesion as 'benign' or 'suspicious for malignancy'." # **Dermoscopy training:** "During the first phase of the study, simple protocols for diagnosis were made up and given to the participants via e- learning courses, direct meetings, and involving self- assessment procedures (Pagnanelli 2003)" Length of training NR **Training format:** Online/ In-person teaching/ Self assessment procedures

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	No
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	No
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus follow up (Reference is expert diagnosis for Teledermatology component of study)

	Histology (not further described): n=16; Disease positive: 5; Disease negative: 11 *Clinical FU (6 months) plus histology of suspicious lesions: n*=219; Disease positive: 0; Disease negative: 208 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 5 Other: 230 benign

Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: NR **Time interval to reference test**: NR **Time interval between index test(s)**: NR

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Yes
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	High risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	Yes
Are there concerns that the test comparison differs from the review question?	Low concern

Notes

Notes	

Guitera 2009a (Modena)

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** Sept 2004 - Aug 2007 **Country:** Italy (and Australia - see Guitera 2009b (Sydney))

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Lesions suspicious of melanoma based on dermatoscopic diagnostic criteria or lesion change; included only a random sample of 50% of benign nevi observed during time period **Setting:** Secondary (general dermatology); Department of Dermatology, University of Modena, Italy **Prior testing:** Clinical and/or dermatoscopic suspicion/Changes on digital monitoring **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** Location/site of lesion lesions on soles/palms excluded; Lentigo maligna excluded; lesions used in previous assessments or RCM model development **Sample size (patients):** No. included: 195 **Sample size (lesions):** No. included: 195 **Participant characteristics:** Median age: 42 (7 to 88yrs); IQR 32, 59; Male: 51.3% **Lesion characteristics:** Pigmented: 92%; 8% amelanotic lesions or those with tan, light gray, or pale blue pigment only). Median thickness 0.65mm (IQ 25, 75: 0.23, 0.98)

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Pattern analysis **Method of diagnosis:** In person diagnosis; at time of first consultation and prior to RCM **Prior test data:** Clinical examination **Diagnostic threshold:** Not reported **Diagnosis based on:** Single observer (n=1) **Observer qualifications:** Dermatologist; Not clearly reported, but is study co-author **Experience in practice:** High experience **Experience with dermoscopy:** High experience; described as Modena expert based in Dermatology Dept

	Other detail: hand-held dermoscope (Delta 10, Heine, Herrsching, Germany).

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described) Disease positive: 79 / Disease negative: 116 **Target condition (Final diagnoses)** Melanoma (invasive): 61; Melanoma (in situ): 18 Benign naevus: 116 (78 compound, 0 dermal, 16 junctional, and 22 Spitz)
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** Only 50% of imaged nevi were included (randomly selected from the image database prior to analysis) to reduce the MM/nevus ratio **Time interval to reference test:** Consecutive; imaged prior to biopsy
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	No
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Guitera 2009b (Sydney)

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective; dermoscopic images assessed remotely from the patient **Period of data collection** Sept 2004 - Aug 2007 **Country** Australia (and Italy - see Guitera 2009a (Modena))
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Lesions suspicious of melanoma based on dermatoscopic diagnostic criteria or lesion change **Setting:** Specialist clinic; Sydney Melanoma Diagnostic Centre, Australia **Prior testing:** Clinical and/or dermatoscopic suspicion/Changes on digital monitoring

	Setting for prior testing: Specialist clinic **Exclusion criteria:** Location/site of lesion lesions on soles/palms excluded. Lentigo maligna excluded; lesions used in previous assessments or RCM model development **Sample size (patients):** No. included: 131 **Sample size (lesions):** No. eligible 156 No. included: 131 **Participant characteristics:** Median age: 52 (19 to 90yrs); IQR 40, 63; Male: 58.8% **Lesion characteristics:** Pigmented: 84%; 16% amelanotic lesions or those with tan, light gray, or pale blue pigment only). Median thickness 0.40mm (IQ 25, 75: 0, 0.84)

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Pattern analysis **Method of diagnosis:** Dermoscopic images **Prior test data:** Lesion site and age available to observer; dermoscopy diagnosis of Sydney lesions was made retrospectively on the images in a random order, blinded to RCM and pathological diagnosis but not to information of site and age, by a Modena expert (GP) using pattern analysis (Pehamberger 1993) **Diagnostic threshold:** Not reported **Diagnosis based on:** Single observer (n=1); **Observer qualifications:** assume Dermatologist; described as Modena expert based in Dermatology Dept **Experience in practice:** High experience **Experience with dermoscopy:** High experience **Other detail:** Sydney - high-resolution digital oil immersion dermoscopy camera (Sentry, Polartechnics Ltd, Sydney, NSW, Australia)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (no further details) Disease positive: 44 / Disease negative: 87 **Target condition (Final diagnoses)** Melanoma (invasive): 26; Melanoma (in situ): 16 Benign naevus: 87 (49 compound, 9 dermal, 26 junctional, and 3 Spitz)
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** 25 lesions out of 156 were rejected for poor quality dermoscopy image, blinded to the diagnostician **Time interval to reference test:** imaged prior to biopsy Time interval between index test(s): N/A
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Haenssle 2010a (FV)

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** 1998-2008 **Country:** Germany

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Participants at increased risk for melanoma: >50 common and/or <=3 atypical nevi; atypical mole syndrome (AMS); or familial atypical mole and multiple melanoma syndrome. [FV - first visit data FU - follow-up data] **Setting:** Secondary (Dermatology) **Prior testing:** All identified as high risk **Setting for prior testing:** NR **Exclusion criteria:** Patients showing melanoma development on pre-existing pigmented lesions during the following 12 months after the analysed time frame **Sample size (patients):** 688 **Sample size (lesions):** 11,137 **Participant characteristics:** Mean age 42 (range NR). 60% male. Group 1 (50 common and/or <= atypical nevi) 67%; Group 2 (AMS) 31.8%; Group III (familial atypical mole and multiple melanoma syndrome) 1.2%. Personal history of melanoma (29.2%); Family history of melanoma (13.1%); High number (>50) of nevi (56.4%). **Lesion characteristics:** NR

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** 7PCL **Method of diagnosis:** In person **Prior test data:** Also considered lesional history (eg, increase in size, itching, scaling, change in color, intermittent bleeding), and the ugly duckling sign (Grob 1998) and 'moles-breed-true' concept (Scope 2006). Lesions scoring <3 on 7PCL were excised if these other factors were present at first visit (FV). Lesions scoring <3 with defined clinical or dermatoscopic criteria of atypia (eg, asymmetry in shape, irregular margin, variegated colour, prominent pigment network) (Ascierto 2000) were marked on digital overview images and electronically stored by using two digital dermatoscopy systems for follow up (FU). **Diagnostic threshold:** >=3 **Diagnosis based on:** Consensus of 2 **Observer qualifications:** Dermatology residents (n=13); supervised by experienced dermatologist **Experience in practice:** NR **Experience with dermoscopy:** High; formally trained in dermoscopy

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and	**Reference standard:** Histology or FU (every 3, 6, or 12 mos)

reference standard(s)	**Target condition (Final diagnoses)** Invasive melanoma 77; Melanoma in *situ* 50; BCC 2 Benign nevi 1047; Spitz nevi 16; Sbhorrheic keratosis 12; Other benign 9935 (not excised)

Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** None reported **Time interval to reference test**: Consecutive

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Haenssle 2010b (FU)

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** 1998-2008 **Country:** Germany

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Participants at increased risk for melanoma: >50 common and/or <=3 atypical nevi; atypical mole syndrome (AMS); or familial atypical mole and multiple melanoma syndrome. [FV - first visit data FU - follow-up data] **Setting:** Secondary (Dermatology) **Prior testing:** All identified as high risk **Setting for prior testing:** NR **Exclusion criteria:** Patients showing melanoma development on pre-existing pigmented lesions during the following 12 months after the analysed time frame **Sample size (patients):** 688 **Sample size (lesions):** 11,137 **Participant characteristics:** Mean age 42 (range NR). 60% male. Mean age 42 (range NR). 60% male. Group 1 (50 common and/or <= atypical nevi) 67%; Group 2 (AMS) 31.8%; Group III (familial atypical mole and multiple melanoma syndrome) 1.2%. Personal history of melanoma (29.2%); Family history of melanoma (13.1%); High number (>50) of nevi (56.4%). **Lesion characteristics:** NR

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** 7PCL **Method of diagnosis:** In person **Prior test data:** Also considered lesional history (eg, increase in size, itching, scaling, change in color, intermittent bleeding), and the ugly duckling sign (Grob 1998) and 'moles-breed-true' concept (Scope 2006). Lesions scoring <3 on 7PCL were excised if these other factors were present at first visit (FV). Lesions scoring <3 with defined clinical or dermatoscopic criteria of atypia (eg, asymmetry in shape, irregular margin, variegated colour, prominent pigment network) (Ascierto 2000) were marked on digital overview images and electronically stored by using two digital dermatoscopy systems for follow up (FU). **Diagnostic threshold:** >=3 **Diagnosis based on:** Consensus of 2 **Observer qualifications:** Dermatology residents (n=13); supervised by experienced dermatologist **Experience in practice:** NR **Experience with dermoscopy:** High; formally trained in dermoscopy

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard:** Histology or FU (every 3, 6, or 12 mos); mean FU 44.28 (range 2-123) months **Target condition (Final diagnoses)** Invasive melanoma 77; Melanoma in *situ* 50; BCC 2 Benign nevi 1047; Spitz nevi 16; Sbhorrheic keratosis 12; Other benign 9935 (not excised)
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** None reported **Time interval to reference test**: Consecutive
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Hauschild 2014

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** RCT of diagnosis based on clinical/dermoscopic images versus same plus MelaFind, with observers randomised between arms. Lesions selected on a case control type basis with cases and controls sampled from a previous study (Monheit 2011). **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR **Country** US
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Subset of pigmented skin lesions evaluated in Monheit 2011; melanoma and non-melanoma randomly selected. **Setting:** Mixed Secondary/Private; Lesions sampled from Monheit trial "Seven clinical sites with 23 investigators participated in this trial. Three sites were academic institutions (University of Pittsburgh, Duke University, and Northwestern University), and 4 sites were dermatologic practices highly experienced in managing PLs." (Monheit 2011) **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Not reported **Exclusion criteria:** Ulcerated or non-pigmented lesions, or located on excluded anatomic sites. Lesions with prebiopsy clinical diagnoses of melanoma were excluded from Monheit 2011 **Sample size (patients):** No. included: 130 **Sample size (lesions):** No. eligible: 1632 lesions in Monheit trial; No. included: 130

	Participant characteristics: None reported **Lesion characteristics:** Head/Neck: 23%; Trunk: 41.5%; Upper limbs/shoulder: 20%; Lower limbs/hip: 16.2%. Median thickness (melanomas) 0.39mm (range 0.12 to 1.2mm)

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Clinical photographs and dermoscopic images (Arm 1 and Arm 3 of trial; Arm 2 included MelaFind images) **Prior test data:** Clinical images (overview and close up); plus 24 items regarding patient demographics and risk factors for melanoma such as: personal or family history of melanoma, number of atypical nevi, Fitzpatrick skin type, number of severe sunburns before and after age 20, etc. **Diagnostic threshold:** Biopsy decision. **Diagnosis based on:** Average. Board-certified dermatologists who were members of a public dermatology list volunteered to participate in the trial. Selection was made on a first-come basis with randomisation between two study arms until at least 65 dermatologists participated in each Arm. Of the 227 dermatologists registered, 211 completed at least 78 cases and therefore were considered eligible. FInally included 101 of 108 dermatologists in Arm 1 and 101 of 108 dermatologists in Arm2 (MelaFind). A third arm included 9 of 12 Pigmented Skin Lesion experts "prospectively identified by the Principal Investigator based on field standing prior to participant recruitment" **Observer qualifications:** Dermatologists **Experience in practice:** High; all board certified, in Arm 1 >90% had more than 10 years experience in practice; Arm 3 consisted of PSL experts **Experience with dermoscopy:** High; for Arm 1 all except 6 were trained in dermoscopy use and 80/101always or almost always used dermoscopy for PSLs; Arm 3 consisted of PSL experts

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: From Monheit 2011 - The electronic case record included details of the "prebiopsy diagnoses (without dermoscopy and, if available, with dermoscopy) by the examining dermatologists", "if the dermatologic diagnosis was not melanoma, the reason for the biopsy was selected from the following: nonmelanoma skin cancer, patient's concern, patient's discomfort, cosmetic, or, if dermoscopic evaluation was used, clinical concern. A histologic specimen with the standard hematoxylin-eosin staining was provided for each lesion." "Histologic slides for each lesion ... were evaluated by 2 independent dermatopathologists. In cases of significant discordance, histologic slides were evaluated independently by a third dermatopathologist. When 1 dermatopathologist diagnosed melanoma and 2 others diagnosed a benign lesion, histologic slides were sent again to the dermatopathologist who diagnosed melanoma for a blind rereview." Disease positive: 65; Disease negative: 65 **Target condition (Final diagnoses)** Melanoma invasive: 36; Melanoma in situ: 29 'Benign' diagnoses: 65
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Appears consecutive
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	

Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Kittler 1998

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Unclear **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** Not reported **Country:** Austria

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Unclear
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesion (PSL) images 'selected' by PSL experts from pigmented lesion image database on the basis of quality of the photograph and the difficulty of diagnosis; all "melanomas selected provided only subtle ELM features as clues to the malignancy of the lesion and were difficult to differentiate from benign PSLs". **Setting:** Secondary (not further specified) **Prior testing:** Dermatoscopic suspicion in all cases **Setting for prior testing:** Secondary (general dermatology); selected from PSL database **Exclusion criteria:** None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. included: 50 **Participant characteristics:** None reported **Lesion characteristics:** median Breslow thickness of the MMs: 0.7mm (IQR 0.5-0.95mm)

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** No algorithm **Method of diagnosis:** Dermoscopic images; both photographic slides and compressed digital images assessed to determine whether compressed images are sufficiently informative for diagnosis; 2x2 based on digital images used for primary analysis **Prior test data:** No further information used. Images viewed in two sessions; in each session 25 slides and 25 digital images were viewed **Diagnostic threshold:** Clinical diagnosis; Rated as definitely or probably melanoma; unclear whether 2x2 based on 'definite' only as test positive or definite/probable combined **Diagnosis based on:** Single observer; n= 8 readers, reported separately. **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Not described; described as 'pre-trained in ELM'

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard:** Histological diagnosis alone (no further details) Disease positive: 23; Disease negative: 27

	Target condition (Final diagnoses) Melanoma (invasive or *in situ*): 23 Sebhorrheic keratosis: 1; Atypical naevus 17; Common naevus 9

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** Poor quality images excluded; "selected' from pigmented lesion image database on the basis of quality of the photograph" **Index test to reference standard interval:** Not reported

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Kittler 1999

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** November 1996 to November 1997 **Country** Austria (from authors' institution)

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions < 1 cm in diameter, consecutively excised **Setting:** Secondary (general dermatology) From authors' institution **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Not reported **Exclusion criteria:** lesion size >=1cm **Sample size (patients):** No. included: 352 **Sample size (lesions):** No. included: 373 **Participant characteristics:** Mean age 52 (SD 17 years); Male: 49% **Lesion characteristics:** median thickness 0.65mm (range, 0.2 to 2 mm).

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** ABCD; ABCDE (developed in this study) **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination **Diagnostic threshold:** Range of numerical thresholds evaluated. 'Standard' ABCD applied as previously described by Stolz 1994 and Nachbar 1994. Sensitivities reported for a range of specificities but cut-offs not reported (author communication suggested a threshold of >4.75 was used but not clear which se/sp pair this relates to); randomly selected dataset at 75%specificity for inclusion in primary analysis. 'Enhanced' ABCD-E algorithm accounts for patient report of changes in the lesion within the previous year. The overall score was calculated by adding 1.2 to the standard ABCD score for changing lesions and subtracting 0.8 from the standard ABCD score for non-changing lesions according to the results of a multivariate analysis. ABCDE results reported at cutoffs ranging from 1.30 to 7.35. **Diagnosis based on:** Unclear; appears to be in clinic diagnoses (n=NR) **Observer qualifications:** Not reported; likely dermatologists **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	No
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	High risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: "After excision all lesions were subjected to standard histopathologic examination. The histologic diagnosis of an atypical nevus was based on the following criteria: cellular atypia, lentiginous hyperplasia of the epidermis, fibroplasia, bridging of rete ridges, suprabasal melanocytes, junctional nest disarray. Disease positive: 73; Disease negative: 283 **Target condition (Final diagnoses)** Melanoma (invasive): 55 (51 superficial spreading,4 nodular, 15 lentigo maligna, 3 otherwise nonclassified melanomas); Melanoma (*in situ*): 18 Sebhorrheic keratosis: 4; 126 (35.4%) common nevi, 113 (31.7%) atypical (dysplastic) nevi, 3 (0.8%) congenital nevi, 13 (3.7%) pigmented Spitz nevi, 7 (0%) blue nevi, 2 (0.6%) combined nevi, 14 (3.9%) solar lentigines, 1 dermatofibroma
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** Non melanocytic lesions (n=17; including angiomatous tumours, pigmented seborrhoeic keratosis, dermatofibromas, and pigmented basal cell carcinomas) easily distinguished by standard ELM criteria and pattern analysis **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Kittler 2001

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** Not reported **Country** Not reported
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Images of naevi from patients with multiple atypical naevi undergoing digital dermoscopy follow-up. All melanomas were excised due to changes on follow up; benign melanocytic skin lesions included were taken at random from the participants with melanoma plus other randomly selected patients with multiple atypical naevi. **Setting:** Secondary (assumed); states "a database" Auth inst: Dept Dermatology, Uni Vienna **Prior testing:** All undergoing follow-up **Setting for prior testing:** Not reported

	Exclusion criteria: None reported **Sample size (patients):** No. eligible: NR; No. included: 20 **Sample size (lesions):** No. eligible: NR; No. included: 80 **Participant characteristics**: None reported **Lesion characteristics**: None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** Unclear **Diagnostic threshold:** Data extracted for excise decision; data also presented for 3 option response of excise/follow-up or no intervention. **Diagnosis based on:** Average (n=24); three groups were recruited according to experience but 2x2 can be extracted only for overall average result, individual group results presented only graphically. **Observer qualifications:** Dermatologist **Experience in practice:** Not reported **Experience with dermoscopy:** Mixed; group 1 (n=9) had basic dermoscopy experience with no formal training, group 2 (n=10) had dermoscopy training but only basic experience with digital dermoscopy, and group 3 included experienced dermatologists trained in dermoscopy and using digital dermoscopy routinely to follow-up melanocytic lesions.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus follow up Details: All lesions were excised (n=20; including all 10 melanomas) or had at least 2 years of follow-up with no morphologic changes during multiple examinations (n=60; all benign) **Target condition (Final diagnoses)** Melanoma (invasive): 5, Melanoma (*in situ)*: 5 Benign melanocytic lesions: 70
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** None reported **Time interval to reference test:** Not reported for histology; Clinical follow up lasted up to 2 years
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Yes
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Krahn 1998

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** NR **Country:** Germany

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Excised pigmented skin lesions **Setting:** Secondary (general dermatology) **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** No. included: 80 **Sample size (lesions):** No. included: 80 **Participant characteristics:** None reported **Lesion characteristics** range in thickness (melanomas) 0.18-1.9mm; 29/39 <0.76mm; 7/39 0.76-1.5mm; 3/39 >1.5mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI):** No algorithm reported **Method of diagnosis:** In person diagnosis **Prior test data:** Unclear **Diagnostic threshold:** Not reported no details **Diagnosis based on:** Single observer (n=1) **Observer qualifications:** Not reported; likely Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Not described # **Dermoscopy** **Method of diagnosis:** In person diagnosis **Prior test data:** Unclear **Diagnostic threshold:** Not reported no details Test observers as described for Visual Inspection (above)

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone including histometrics Disease positive: 39; Disease negative: 41 **Target condition (Final diagnoses)** Melanoma (invasive): 39 (SSM, lentigo MM, nodular M) Benign naevus: 37 common nevus; 3 dysplastic nevus, 1 Spitz nevus
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: none reported **Time interval to reference test:** not reported **Time interval between index test(s)**: not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	tbc
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	High risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	Yes
Are there concerns that the test comparison differs from the review question?	Low concern

Notes

Notes	

Kreusch 1992

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** NR; 1.5 year period **Country:** Germany
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented lesions suspected to be malignant melanoma with adequate photo- documentation and histology results **Setting:** Secondary (dermatology) **Prior testing:** NR **Setting for prior testing:** NR **Exclusion criteria:** Non-melanocytic lesions **Sample size (patients):** Total 856; NR for final sample **Sample size (lesions):** 265 melanocytic/1506 lesions included (317 excised and 52 NML excluded) **Participant characteristics:** NR **Lesion characteristics:** NR
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** Algorithm from Kreusch 1991 **Method of diagnosis:** Image based **Prior test data:** None; slides labelled only with patient code and lesion localisation **Diagnostic threshold:** >=9; scored diameter >5mm; border irregularity; loss of surface's microstructure; scaling/erosion/ulcer; capillaries (each 1 point); multicomponent architecture; greyish colour (each 3 points) melanophages (6 points); pseduopods (10 points); regression (10 points)

	Diagnosis based on: Single observer **Observer qualifications:** Dermatologist (assumed) (n=1; 'experienced') [also presents results for inexperienced student – data not included] **Experience in practice:** 'experienced' **Experience with dermoscopy:** 'experienced'

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histology **Target condition (Final diagnoses)** Invasive melanoma 96; benign nevi 169
Is the reference standards likely to correctly classify the target condition?	Unclear
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** *52 NML excluded from second step evaluation* **Time interval to reference test**: NR
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Langley 2007

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** February 2002 to May 2005 **Country:** Canada
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients with suspicious pigmented lesions scheduled for biopsy due to clinical suspicion of malignancy determined by clinical appearance or a history of change in the lesion. **Setting:** Specialist unit (skin cancer/pigmented lesions clinic); Division of Dermatology Pigmented Lesion Clinic and the Plastic Surgery Clinics

	Prior testing: Clinical suspicion of malignancy **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** Non-pigmented; Physically inaccessible lesion site; previous diagnostic biopsy of the lesion. **Sample size (patients):** No. eligible: 127; No. included: 125 **Sample size (lesions):** No. eligible: 127; No. included: **Participant characteristics:** Mean age 44.2 yrs, range 16 to 84 yrs **Lesion characteristics:** median thickness 0.62 mm, range 0.20 mm to 7.92 mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** Pattern analysis **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** Pattern analysis; diagnosis of melanoma **Diagnosis based on:** Single observer (n=1) **Observer qualifications:** Not reported likely dermatologist; "Clinical, dermoscopic and confocal examinations were conducted sequentially by a single reviewer" and a diagnosis recorded after each. **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: "When CSLM imaging was complete, the lesions were removed by excisional biopsy. A definitive diagnosis was made by a dermatopathologist with conventional hematoxylin-eosin stained histopathological sections." Disease positive: 37; Disease negative: 88 **Target condition (Final diagnoses)** Melanoma (invasive): 22; Melanoma (in situ): 15 Benign naevus: 88
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** Two patients were excluded from the data- base due to technical difficulties with the imaging. **Index test to reference standard interval:** When CSLM imaging was complete, the lesions were removed by excisional biopsy
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias

Comparative	

B. Concerns regarding applicability

Notes

Notes	

Lorentzen 1999

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** Between 1994 and 1997 **Country:** Denmark

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients with lesions suspicious for CMM referred to outpatients clinic; only excised included **Setting:** Not reported **Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion **Setting for prior testing:** Not reported **Exclusion criteria:** Poor quality index test image (considered under flow/timing) **Sample size (patients):** No. eligible: 242; No. included: 232 **Sample size (lesions):** No. eligible: 242; No. included: 232* **Participant characteristics:** None reported **Lesion characteristics:** None reported *NB Not all cases were assessed by all observers; 2x2 are based on presented sensitivity and specificity estimates for full dataset of lesions; "the dermatoscopy experts assessed almost all cases (98 ± 100%), whereas the non-expert group completed fewer assessments, from 76 to 98%.

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** Clinical photographs **Prior test data:** No further information used; no option to change clinical diagnosis after viewing dermoscopic image **Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone; clinical images presented before dermoscopic images **Diagnostic threshold:** Not reported; clinical diagnosis **Diagnosis based on:** Average; n= 9 **Observer qualifications:** Dermatologist **Experience in practice:** High; Moderate; Mixed (average reported); 4 'experienced dermatologists' (4-5 years daily experience) & 5 'non-expert dermatology residents' (1-2 years interest and formal training in dermatoscopy] **Experience with index test:** High; Moderate; Mixed # **Dermoscopy:** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** Clinical image presented first **Diagnostic threshold:** Clinical diagnosis"; observers were familiar with both the ABCD-rule of dermatoscopy proposed by Stolz et al. (6) and Kenet et al.'s risk stratifying algorithm of pigment network features of dermatoscopy (8). The observers were not constrained by either of the rules. The ABCD scores were not used to obtain the diagnoses. Rather a pattern recognition process was intended." # **Dermoscopy training:** described as "formal training" **Training format:** Non experts had undergone prior training in dermoscopy (not documented)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes

Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: a co-author from Dept of Pathology "re-evaluated all cases to confirm the pathology diagnosis, which was used as the gold standard in this study." Disease positive: 65 ; Disease negative: 167 **Target condition (Final diagnoses)** Melanoma (invasive): 49 'malignant melanoma' BCC: 16 Sebhorrheic keratosis: 12; Benign naevus: 137 (pigmented nevi=116; blue nevi=16; atypical nevi=5); Other: 18 (spitz nevi, Bowen's disease, sarcoid, nevus spilus, hemangioma, and others)
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: 10 cases were "considered unfit for evaluation" due to poor quality image **Reference interval**: "biopsy specimens...were obtained after the clinical and dermatoscopic photographs had been performed"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	tbc
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	Yes
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Lorentzen 2000

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** 1995-1999 **Country** Not clear; authors from Denmark and US
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions from patients consecutively referred to the skin cancer outpatient clinic with available clinical photographs, dermatophotographs and a subsequent excision biopsy were included **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** None reported **Sample size (patients):** No. included: 258 **Sample size (lesions):** No. included: 258

	Participant characteristics: None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** ABCD; Kenet Risk Stratification **Method of diagnosis:** Dermoscopic images; "Slides were projected to an 80 x 120 cm screen in a darkened room. Based on time studies in the outpatient clinic, each patient case was shown for approximately 3 min... additional time was allowed if any needed it." **Prior test data:** Cliincal photographs also projected **Diagnostic threshold:** ABCD - 'possible' MM: >4.75; 'probable' MM: >5.45. Risk stratification method: 'possible' MM: stratum 1 or 2; 'probable' MM: stratum 1 only (1: Probable CMM: Pseudopods; Radial streaming; Heterogeneity of PN with thick dark extensions at the edge; Blue-grey areas, white scar like areas and presence of PN2: Possible CMM: Marked irregular network with irregular pigment confluence) **Diagnosis based on:** Single observer (n=3; performed independently) **Observer qualifications:** Dermatologist **Experience in practice:** High; Senior dermatologists - "Three senior dermatologists with more than 5 years daily experience in clinical use of dermatoscopy and familiar with (both) dermatoscopic (algorithms)" **Experience with dermoscopy:** High; > 5 years each

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone - Details: Lesions underwent HE-staining, as well as HMB-45 and S-100 immunostaining to identify melanocytic lesions. Breslow depth and Clark level were determined. All cases were assessed by an experienced dermatopathologist. Disease positive: 64; Disease negative: 194 **Target condition (Final diagnoses)** Melanoma (invasive): 64 CMM BCC: 25 Sebhorrheic keratosis: 14 Benign naevus: 135; Dysplastic naevus 3; Other: 11 blue naevi, 1 pigmented Spitz naevus, plus one each of were angioma, haemorraghia, papilloma and dermatofibroma.
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Time interval to reference test:** Appears consecutive; "Only patients having taken clinical photographs, dermatophotographs and a subsequent excision biopsy were included"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Lorentzen 2008

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Unclear **Period of data collection** not reported **Country** Denmark

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients referred to the specialist naevus clinic for lesion excision **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria: N**ot specified **Sample size (patients):** No. eligible: 120; No. included: 119 **Sample size (lesions):** No. included: 119 **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** Mixed/no algorithm; describes using "the risk stratification and pattern analysis procedure as described by Kenet 2001 and Lorentzen 2000". **Method of diagnosis:** Dermoscopic images; compared accuracy using standard dermoscopy images (Dermaphot) and images obtained using a globe magnifier. Slides were randomised and evaluated on 2 different occasions with 3 week intervals **Prior test data:** No further information used **Diagnostic threshold:** Observer correct diagnosis of each lesion type **Diagnosis based on:** Unclear (assumed Average) (n=NR) **Observer qualifications:** Dermatologist **Experience in practice:** High; "dermatologists who have performed dermatoscopy for 5–10 years, published scientific papers on dermatoscopy and carried out pre- and post specialist training in dermatoscopy" **Experience with dermoscopy:** High

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard:** Histological diagnosis alone Details: used haematoxylin-eosin staining a well as histochemistry was performed using S-100 and HMB-45 on suspect melanoma lesions. Disease positive: 24; Disease negative: 95 **Target condition (Final diagnoses)** Melanoma (invasive): 24 BCC: 13

	Mild//moderate dysplasia: 2; Sebhorrheic keratosis: 9; Haemangioma: 2; Naevus pigmentosus- 69

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** One dermatofibroma excluded **Time interval to reference test:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Malvehy 2014

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective; dermoscopic images assessed remotely from the patient **Period of data collection:** March 2010 and November 2011 **Country:** conducted at five American and 17 European investigational sites (Sweden, Germany, Austria, Hungary, U.K. and Spain);
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** All patients with skin lesions selected for total excision to rule out melanoma; dermatologists were encouraged to enrol a mix of lesions with an even distribution of low-, medium and high-risk lesions. **Setting:** Secondary; authors institutions primarily listed as Dept Dermatology with one "Dermatology Clinical Research Center" **Prior testing:** Selected for excision **Exclusion criteria:** lesions < 2 mm or > 20 mm and those located: on acral skin, e.g. sole or palm; areas of scars, crusts, psoriasis, eczema or similar skin conditions; hair-covered areas, e.g. scalp, beards, moustaches or whiskers; genitalia; in an area that has been previously biopsied or subjected to any kind of surgical intervention or trauma; mucosal surfaces; with foreign matter, e.g. tattoo or splinter; acute sunburn; or skin surface not measurable, e.g. lesion on a stalk; surface not accessible, e.g. inside ears, under nails or not intact (measurement area), **Sample size (patients):** No. eligible: 1951; No. included: 1611 for Nevisense and NR for visual inspection and dermoscopy **Sample size (lesions):** No. eligible: 2416; No. included: 1943 for Nevisense and 1701 for visual inspection and dermoscopy **Participant characteristics:** For Nevisense sample: median age: 48y (range 18 to 91); male 47.5%; 97.5% of white ethnicity. Fitzpatrick skin types: I (7.3%); II (48.6%); III (37%); IV (9.8%); V (1.4%); VI (0.1%) **Lesion characteristics:** median Breslow thickness of 0.57 mm (153 invasive melanomas);
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy**; ABCD; 7-point checklist; revised 7-point checklist; overall diagnosis (methods describe evaluation of the clinical ABCD rule but results not presented in Table) **Method of diagnosis:** Image-based; "A photograph and dermoscopic image of each included lesion was taken before and after Nevisense measurements" **Prior test data available:** Clinical and dermoscopic images presented together; observers were blinded to Nevisense result **Diagnostic threshold:** ABCD - >4.75 and >5.45; 7-point checklist and revised 7-point checklist – NR, referenced to Argenziano 1998; overall diagnosis based on grading (0 to 10) on a visual classification board with a fixed cut-off at 4. **Diagnosis based on:** Unclear; (n=3) **Observer qualifications:** Dermatologists; "images were reviewed by three dermatologists with 2–5 years of experience in dermoscopy assessment. The option to reach out to additional experienced dermoscopists in difficult cases was allowed" **Experience in practice:** Not described **Experience with dermoscopy:** High; 2-5 years

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	No
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Type of reference standard:** Histological diagnosis alone **Details:** Lesions were excised and underwent usual histopathology at investigational site. A further histopathological evaluation was undertaken for study purposes by a panel of three experienced histopathologists who evaluated each lesion independently; blinded from the investigational site's original histopathology diagnosis. If they agreed, the diagnosis was considered as the histopathological gold standard (HGS); if there was significant disagreement regarding malignancy the slides were submitted to two additional experts whose diagnosis was then chosen as the HGS if they reached agreement. In case of disagreement by the two additional reviewers, the corresponding lesion was excluded from the efficacy analysis. Disease positive: 238 for VI/dermoscopy; Disease negative: 1440 **Target condition (Final diagnoses)** For VI/Dermoscopy sample – 238 melanomas including 112 *in situ* Breakdon of non-diseased not provided for VI/dermoscopy sample For Nevisense sample (includes additional 242 lesions: 153 invasive melanomas, 112 melanoma in situ, 48 BCC, 1 invasive SCC; 1 Merkel cell carcinoma 157 severely dysplastic, 988 mild to moderate dysplasia, 352 benign nevi, 5 spitz nevi, 51 seborrheic keratosis, 6 SCC in situ; 8 AK; 61 other
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** 473 excluded from Nevisense analysis; all reasons listed; primary reason was investigator oversight or the inability to render a final histopathological diagnosis; 74 exclusions were device-related (60 with inadequate reference measurement quality and 14 to device failure). A further 242 were excluded from VI/Derm analysis due to image quality (12% of visual inspection/dermoscopy sample) **Index test to reference standard interval:** Appears consecutive; prospective recruitment with imaging and then "eligible and evaluable lesions were excised and subjected to the investigational site's histopathology evaluation and managed accordingly." "A postprocedure follow-up either by a telephone call or at a participant's visit to the investigational site was conducted at 7 +/- 3 days after the Nevisense evaluation, at which time the patient was evaluated for any adverse events."
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	**Interval between index tests** Consecutive; "A photograph and dermoscopic image of each included lesion was taken before and after Nevisense measurements to document evaluation according to the protocol."

B. Concerns regarding applicability

Notes

Notes	

Menzies 1996

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Unclear. Abstract describes including a random sample of excised lesions from a larger database. **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR **Country** Australia **Test set derived** NR; describes 'division' into a training set and a test set.
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Unclear

Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions from the Sydney Melanoma Unit with dermoscopic images and histological diagnoses; melanomas and randomly selected clinically atypical nonmelanoma lesions were included. **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Selected for excision **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** Unequivocal nonmelanoma excluded **Sample size (patients):** No. included: NR **Sample size (lesions):** No. included: 385 **Participant characteristics:** None reported **Lesion characteristics:** None reported
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Menzies criteria **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used **Diagnostic threshold:** Presence of 2 negative features and at least one positive feature. Negative features: point and axial symmetry of pigmentation or presence of only a single colour. Positive features of melanoma: multiple (5-6) colors; blue-white veil; multiple brown dots; multiple blue/gray; peripheral black dots or globules; a broadened network; pseudopods; radial streaming; scarlike. **Diagnosis based on:** Unclear (n=NR) **Observer qualifications:** Not reported; likely dermatologists **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described) Disease positive: 107; Disease negative: 278 **Target condition (Final diagnoses)** Melanoma (invasive): 107; BCC: 18 Ephelis/lentigo 17; Sebhorrheic keratosis: 23; Benign acquired nevi - 58; Dysplastic nevi - 105; Blue nevi 11; Spiz nevi 6; spindle cell nevus 2; dermatofibroma 2; hemangioma 13; solar keratosis 9; other 14
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes

If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Menzies 2005

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** June 1998 to September 2003 **Country** Multicentre (Australia, US, Germany)

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** All melanocytic lesions from the independent test set taken at the Sydney Melanoma Unit that had clinical and dermoscopy photographic images; lesions imaged prior to excision due to clinical suspicion of malignancy or because of short-term digital monitoring (study was part of a larger multicentre study of SolarScan). **Setting:** Specialist unit **Prior testing:** Clinical suspicion of malignancy or requirement for short-term digitial monitoring **Setting for prior testing:** Specialist unit **Exclusion criteria:** Awkwardly situated lesions (eg, eyelids, some parts of the pinna, some genital sites, and perianal and mucosal surfaces); acral lesions; non-pigmented pure amelanotic lesions (based on dermoscopy imaging); ulcerated lesions, or diagnosed as pigmented basal cell carcinoma, pigmented Bowen disease, or squamous cell carcinoma **Sample size (patients):** No. included: NR **Sample size (lesions):** No. included for Dermoscopy review: 78 (For full SolarScan study - No. eligible: 2430/ No. included: 1644 training; 786 test set) **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** Clinical photographs and patient histories (including details of age, sex, and lesion site; and a recorded history of whether the lesion had, within the past 2 years, bled without being scratched, changed in color or pattern, or increased in size). **Diagnostic threshold:** Data can be extracted at two thresholds:- correct diagnosis of melanoma (in situ or invasive) and excise decision; No details on lesion characteristics used. **Diagnosis based on:** Average according to qualification level (n=13) **Observer qualifications:** GP 3; Dermatology registrar 3; Dermatologists 4; plus 3 international dermoscopy experts who headed pigmented lesion clinics **Experience in practice:** Not described **Experience with dermoscopy:** Expert/High/Moderate/Low

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus other (Full sample n=2430) *Histology:* 71% of full SolarScan study sample including training and test set (n=1725) *Clinical FU plus histology of suspicious lesions;* Length of FU: 3 mo. 26% of full SolarScan study sample (n=632) *Expert opinion.* 3% of full SolarScan study sample were nonmelanocytic pigmented lesions that were diagnosed clinically but not excised (n=73) **Target condition (Final diagnoses).** All numbers are for Sydney Melanoma Unit test sample lesions only (n=78) Melanoma (invasive): 5; Melanoma (in situ): 6; Lentigo maligna: 2 Benign melanocytic lesions: 65
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	No
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	High

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** Poor quality index test image as exclusion criterion - lesions outside the field of view (24x18 mm), contamination of calibration surfaces, or excess artifacts (hair, air bubbles, or movement artifacts). **Index test to reference standard interval: Not described.**
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Menzies 2008

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series? **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** NR **Country:** Multicentre
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Dermoscopic amelanotic (with no melanin pigmentation) or hypomelanotic (a melanin pigmentation area of less than 25% of the total surface area or slightly pigmented but with no dark brown, deep blue, or black pigmentation) lesions. All melanomas included, and a random selection of melanocytic and nonmelanocytic lesions on a non-melanoma to melanoma ratio of 3:1. **Setting:** Multicentre **Prior testing:** NR **Setting for prior testing:** NR **Exclusion criteria:** Lesions were excluded because of poor image quality or because they did not fit within any of the defined pigmentation categories **Sample size (patients):** NR **Sample size (lesions):** 497 **Participant characteristics:** NR **Lesion characteristics:** NR
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** 7PCL; Menzies; 3PCL [new algorithm for distinguishing melanoma from nonmelanoma and any malignant from benign lesions was also developed on 80% of sample and tested on 20% but numbers Disease positive and negative for the test set were not reported to allow 2x2 to be estimated.] **Method of diagnosis:** Image-based **Prior test data:** NR **Diagnostic threshold:** >=3; Menzies standard threshold; >=2 **Diagnosis based on:** Single observer

	Observer qualifications: Dermatologist (assumed) (n=12); clinicians experienced in dermoscopic evaluation scored 99 individual morphological features in approximately equal sample sizes. **Experience in practice:** NR **Experience with dermoscopy:** High

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	No
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histology and FU (no.s NR; some nevi included that showed no changes following consecutive digital monitoring) **Target condition (Final diagnoses)** Invasive melanoma 91; Melanoma *in situ* 14; BCC 126 BCC; cSCC 4 Benign nevi 159; Spitz nevi 11; Sebhorrheic keratosis 22; dermatofibroma 17; Bowen's disease 7; Keratoacanthoma 1; actinic keratosis 8; other 37
Is the reference standards likely to correctly classify the target condition?	Unclear
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** None reported **Time interval to reference test**: NR
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Menzies 2009

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** December 2005 to August 2006 **Country** Australia
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented lesions which, after routine naked eye examination by the GP, would have been biopsied or referred, i.e. a SPL (suspicious pigmented lesion). GPs were recruited from practices with at least 3 clinicians; excluded if they already used dermoscopy or SDDI in their routine practice.

	Setting: Primary **Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion **Setting for prior testing:** Primary **Exclusion criteria:** None reported **Sample size (patients):** NR **Sample size (lesions):** No. included: 374 **Participant characteristics:** None reported **Lesion characteristics** None reported

Are the included patients and chosen study setting appropriate?	Yes
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	Unclear

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Diagnostic threshold:** Not reported. Initial diagnosis recorded along with confidence of diagnosis (scale 1 to 10; 1 not at all confident and 10 extremely confident), certainty of melanoma (scale 0 to 100%; 0 definitely not melanoma and 100 definitely melanoma) and management (biopsy, referral). **Diagnosis based on:** Single observer (n=63; 102 GPs initially recruited; 74 (75%) completed the educational intervention and online assessment; 63 GPs from 19 practices finally participated) **Observer qualifications:** GP **Experience in practice:** Not described **Experience with dermoscopy:** Not fully described; classified as 'trained'. GPs must have each excised or referred >=10 PSL in previous 12-mo period; excluded if dermoscopy or SDDI already used in routine practice. # **Dermoscopy** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes. **Diagnostic threshold:** Not reported. After clinical exam and dermoscopy GP recorded the site of the lesion and the initial diagnosis, confidence of diagnosis, certainty of melanoma and management (as for VI above). Approach to dermoscopy interpretation not further reported; 2x2 can be constructed for decision to Excise or to Excise or Monitor. Triage management options included: biopsy due to clinician concern; biopsy due to patient concern; referral due to clinician concern; referral due to patient concern; short-term SDDI; and patient to return if changes occur. **Test observers** as described for Visual Inspection (above) # **Dermoscopy training:** Online textbook in dermoscopic diagnosis and the use of SDDI, a CD-rom tutorial showing examples of changed and unchanged monitored lesions; a 2-h workshop on the use of the diagnostic devices and recruitment procedures. Assessment through online pre- and post-education intervention test of 245 lesions not seen in the textbook or the CD-rom. Before formal patient recruitment began, GPs assessed at least one pretrial lesion to determine the quality of imaging with the SDDI instrument and undertake completion of trial paperwork. GPs were allowed to practise using the dermoscopy device during this pretrial phase. The pretrial phase of education and run-in period occurred from May 2005 to January 2006. **Length of training:** Self learning plus 2-h workshop **Post-training experience:** 6-12 months **Training format** Online; CD-rom: workshop.

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	No
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	

Target condition and reference standard(s)	**Reference standard -** Histological diagnosis plus other [author confirmed that all melanoma had histological diagnosis and >50% of benign had histology or follow-up] *Histology:* described as conducted to standard practice and not necessarily blinded to the GP's diagnosis.Total excised or referred - 163. Immediate excision/referral - 110. Excision/referral after SDDI - 48. Excision/examination after patient self referral - 5 Disease positive: 37; Disease negative: total of 126 benign or unknown were 'excised OR referred' *Clinical FU plus histology of suspicious lesions:* Short term digital monitoring (SDDI) available as an option for lesions considered not to be melanoma but that were still considered suspicious; follow-up imaging occurred initially at 3 months with any morphological changes to result in biopsy or referral; some lesions continued SDDI for a further 3 months; Length of FU: 3-6 months No. patients: Initially recommended for SDDI: 192; SDDI continued for further 3 months: 6; Underwent SDDI only (no excision): 146 Disease positive: 15 (SDDI then histologically confirmed); Disease negative: 176 benign (incl 1 missed *in situ* melanoma); 4 unknown *Expert opinion:* GPs could refer for specialist opinion or lesions could undergo dermoscopy telemedicine (images reviewed by an expert in dermoscopy and SDDI). Dermoscopy telemedicine was blinded to the GP's diagnosis. Observe for change group, i.e. discharged after dermoscopy: 72 (lus a proportion of those in Excise/refer group will have had expert dx alone but details not given) Disease positive: 0; Disease negative: 71 benign; 1 unknown **Target condition (Final diagnoses)** Melanoma (invasive): 33; Melanoma (in situ): 1 BCC: 6 2 Bowen's disease; 323 benign; 9 unknown

Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	No
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	No
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	High

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: nine lesions with unknown diagnoses, plus BCC and Bowen's excluded from some analyses **Time interval to reference test**: Not reported; Histopathological and specialist examination occurred according to standard practice.

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Yes
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Menzies 2013

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series? **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** NR **Country:** Multicentre (photographic libraries at various institutions; obtained from members of the International Dermoscopy Society from 5 continents)

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Nodular malignant melanoma (an invasive melanoma without an in situ (junctional) component beyond 3 rete ridges of the dermal invasive component) and a random selection of non-nodular invasive primary melanoma, benign nodular melanocytic lesions, and nodular nonmelanocytic lesions at a ratio of NM to other subgroups of 1:2. Nodular benign melanocytic lesions and nodular nonmelanocytic lesions were identified by the clinical appearance of a solitary nodule and confirmed using dermoscopic examination. **Setting:** Mixed **Prior testing:** NR **Setting for prior testing:** NR **Exclusion criteria:** excluded if the image quality was poor **Sample size (patients):** NR **Sample size (lesions):** 467 **Participant characteristics:** excluded if the image quality was poor **Lesion characteristics:** Pigmented 314/467; 67%.

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** ABCD; Menzies; CASH; 7PCL; 3PCL; Menzies algorithm for amelanotic lesions (Menzies 2008) **Method of diagnosis**: Image-based **Prior test data:** NR

	Diagnostic threshold: >5.45; standard Menzies; >=8; standard 7PCL; standard 3PCL; >=1 and >=0 **Diagnosis based on:** Single **Observer qualifications:** Dermatologist (n=2; exp NR) Twelve scorers blinded to the lesion diagnosis scored 99 individual features in each lesion of approximately equal sample sizes, as previously described. **Experience in practice:** NR **Experience with dermoscopy:** High

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	No
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histology or FU ('some' benign melanocytic nevi showed no change over time compared with baseline photographs). **Target condition (Final diagnoses)** Invasive melanoma 217 (incl 83 nodular) Benign naevi 115
Is the reference standards likely to correctly classify the target condition?	Unclear
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** None reported **Time interval to reference test**: NR
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Morales Callaghan 2008

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** 1st January 2005 - 31st December 2005 **Country** Spain
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability

Patient characteristics and setting	**Inclusion criteria:** Randomly selected melanocytic lesions; melanocytic on both clinical and dermoscopic criteria **Setting:** Secondary (general dermatology) **Prior testing:** Dermatoscopic suspicion in all cases **Setting for prior testing:** Not reported **Exclusion criteria:** Palms, soles, mucous membranes of face, under nails; non-melanocytic appearance **Sample size (patients):** No. included: 166 **Sample size (lesions):** No. included: 200 **Participant characteristics:** Mean age 33.7y (SD 14.5), range 8 to 84yrs; Male gender: 64 (38.6%); Fitzpatrick phototype II (44%); type III (41.5%) **Lesion characteristics:** Macular component=181 (90.5%), Papular component=125 (65%) Both = 106 (53%), either one or other = 94 (47%). Asymmetrical 144 (72%). Irregular borders 154 (77%). 4 colours in 40 (20%), 3 colours in 96 (48%), 2 colours in 57 (28.5%), 1 colour in 1 (0.5%). History of bleeding 7 (3.5%). changes reported by patient 154 (77%). Lesion site: trunk 155 (77.5%), including the back in 106 (53%). Lesion size: mean long axis diameter 7.9mm (SD 8.6)mm, mean short axis diameter 5.1 (SD 5).

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes **Other test data**: Appears that dermoscopy was undertaken by same clinician(s) subsequent to clinical evaluation; clinical history was constructed following a standardized protocol and a presumptive clinical diagnosis recorded. Each lesion was then photographed and immediately afterwards examined using a manual dermatoscope **Diagnostic threshold:** Not reported; presumptive clinical diagnosis **Diagnosis based on:** Consensus (n=2) **Observer qualifications:** Dermatologist **Experience in practice:** Not clearly described; assumed to be High - “both dermatologists had experience in dermoscopy.” **Experience with dermoscopy:** Not clearly described; assumed to be High - “both dermatologists had experience in dermoscopy.” # **Dermoscopy** Pattern analysis **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** Not reported; diagnosed "on the basis of predominant dermoscopic pattern(s) using the pattern analysis algorithm" **Test observers** as described for Visual Inspection (above)

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: Lesions described using terminology proposed by US National Insts of Health Disease positive: 6/6 lesions; Disease negative: 194/194 lesions (assuming the 9 'Other' diagnosis lesions were not malignant), or 185/185 (removing the 9 'other' diagnosis lesions from dataset) **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 6 (3%) Other: Atypical mole (104), Common mole (70), congenital nevus (6), Blue nevus (3), Spitz/Reed nevus (1), Spilus nevus (1), Others [unclear whether benign or malignant] (9)

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Exclusions**: none reported **Time interval to reference test**: "Samples for histologic analysis were taken immediately after clinical and dermoscopic examination"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	**Time interval between index test(s)**: Images taken at same time
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	High risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Nachbar 1994

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection** November 1991 to July 1992 **Country** NR (authors institutions Germany and US)
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented melanocytic skin lesions consecutively excised **Setting:** Secondary (general dermatology) **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Not reported **Exclusion criteria:** Unequivocal appearance/diagnosis criteria used to exclude nonmelanocytic described in detail in Table 1; **Sample size (patients):** NR **Sample size (lesions):** No. included: 194 **Participant characteristics:** None reported **Lesion characteristics:** Thickness - 35/69 MM <=0.75mm (50.7%)
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** ABCD **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** >5.45 (determined based on retrospective analysis of the data) For the calculation of ABCD score the criteria of asymmetry (A), abrupt cutoff of the pigment pattern at the border (B), different colors (C), and different structural components (D) were assessed to yield a semiquantitative score (all described in detail). "The results of the retrospective study showed that melanocytic pigmented skin lesions could be differentiated into two diagnostic groups as follows: melanocytic nevi (MN) if the final score was less than 5.45 and MM if the score was higher than 5.45. Retrospective analysis showed an early melanoma could not be completely excluded in all lesions with an ABCD score between 4.75 and 5.45.Therefore these lesions were excised. All lesions were examined by two independent dermatopathologists. **Diagnosis based on:** Unclear (n=NR) **Observer qualifications:** Not reported; Presumably dermatologists; 'colleagues in our department' **Experience in practice:** High experience or 'Expert' **Experience with dermoscopy:** High experience /'Expert' users Study also presents 2x2 data for visual inspection; excluded from review as clinicians 'mostly' also used dermoscope for diagnosis. From text: "In comparing the clinical with the dermatoscopic diagnosis with the ABCD rule it must be noted that all our colleagues in this department referring patients for the study were experienced and in most cases used the dermatoscope without applying the new ABCD rule. Thus clinical diagnosis in our study was expected to be already biased by the dermatoscopic feature and therefore to be more accurate than by the naked eye"

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	No
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	High risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** - Histological diagnosis alone *Histology (not further described)* 194 Disease positive: 69 Disease negative: 125 **Target condition (Final diagnoses)** TARGET CONDITION (Final diagnoses) Melanoma (in situ and invasive, or not reported): 69 BCC: 3 Sebhorrheic keratosis: 19 'Benign' diagnoses: 103 melanocytic naevus
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	Time interval to reference test: NR; Time interval between index test(s): appears consecutive
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Nilles 1994

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** 1989 to 1991 **Country** Germany **Derivation of test set:** Images collected 1989-1990 were used to develop a new algorithm; lesions investigated in 1991 were used for model validation (latter data included in review)
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear

Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic skin lesions that underwent excision **Setting:** Secondary (general dermatology) **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Not reported **Exclusion criteria: N**on-melanocytic appearance **Sample size (patients):** No. included: 260 (1989 to 1990 group); NR for 1991 group **Sample size (lesions):** No. included: 260 (1989 to 1990 group); 209 for 1991 group **Participant characteristics:** None reported **Lesion characteristics** None reported
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** New algorithm **Method of diagnosis:** For training set dermoscopic images were projected onto a screen; method NR for test set (assumed same procedures followed) **Prior test data:** No further information used **Diagnostic threshold:** Significance of '8 clues of malignancy' (ref Braun-Falco 1990) were investigated in data collected 1989-1990. A subset of relevant components were identified and evaluated on the test set of lesions (appears to be presence of any one considered test positive): asymmetrical pigment distribution, more than three colours, asymmetrical depigmentation, black pigment, sharp pigment border and atypical radial streaming **Diagnosis based on:** Single observer (n=1) **Observer qualifications:** NR, likely Dermatologist ('one of the authors') **Experience in practice:** Not described **Experience with dermoscopy:** Not described **Derivation aspect:** The 8 clues of malignancy were graded from 0 (absent) to 3 (distinct) on the test set of lesions (including asymmetrical pigment distribution, more than three colours, black-brown pigment, dark brown pigment, prominent pigment network, asymmetrical depigmentation, peripheral stripes, sharp pigment border and atypical radial streaming). Stepwise logistical regression used to select the variables that resulted int he best model for identification of melanoma

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Unclear
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described) Disease positive: 41 in test set; Disease negative: 168 in test set **Target condition (Final diagnoses)** Full breakdown reported only for training set; for test set: Melanoma (invasive): 41 Benign naevus: 168
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear

Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Pagnanelli 2003

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Unclear; likely a case control type selection process **Data collection:** Retrospective image selection / Prospective interpretation (Dermoscopy training study) **Period of data collection** NR **Country** Italy
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Unclear
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Images of pigmented skin lesions from the training set of the Consensus Net Meeting on Dermoscopy (CNMD) (referenced to Soyer 2001), selected by two experts **Setting:** Unclear **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. included: 20 **Participant characteristics:** None reported **Lesion characteristics:** None reported
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** Pattern analysis; 7-point checklist; ABCD; Menzies criteria **Method of diagnosis:** Clinical photographs and dermoscopic images. Partipants were given a CD with lesion images and asked to evaluate the 20 cases independently over a 20 day period. This was repeated approximately 5 weeks post-dermoscopy training. **Prior test data:** Case notes; "Each case contained the following clinical information: age, sex, skin phototype, total number of naevi, personal and/or family history of melanoma, location, diameter and duration of the lesion, as well as medical history concerning morphological changes within the year preceding excision of the lesion." It appears as so though this information was given to participants along with lesion images. **Diagnostic threshold:** Clinical diagnosis of melanoma; "For each case, the participants completed an electronic data sheet that listed criteria for diagnosing PSLs by pattern analysis and by the various algorithms. Participants offered a dermoscopic diagnosis for each case" **Diagnosis based on:** Average (n= 16); authors' colleagues from dept Dermatology were recruited to participate **Observer qualifications:** Dermatology registrar 9; Dermatologist 4; Medical students 3. **Experience in practice:** Not described **Experience with dermoscopy:** Low; Dermoscopic knowledge of this group consisted only of limited personal experience; none had formal training in this technique and/or used dermoscopy in daily professional practice # **Dermoscopy training:** A one hour lecture introduced the principles of dermoscopy and the algorithms to be evaluated. A Web-based tutorial was then made available and participants were asked to spend one hour per day for two weeks to learn and improve dermoscopy knowledge (http://www.dermoscopy.org) **Training format:** In person and online **Post-training experience:** none reported

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes

If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	No
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	No
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described) Disease positive: 6 (30%); Disease negative: 14 (70%) **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 6 (30%) BCC: 2 (10%) Sebhorrheic keratosis: 2 (10%); Clark nevi 8 (40%); Reed/Spitz nevi 2 (10%)
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Unclear
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Piccolo 2000

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Prospective **Period of data collection:** states 3 months but no specific dates given **Country:** Austria (Graz)
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Lesions included in the study were selected because of their diagnostic difficulty and were excised for a histopathological evaluation. **Setting:** Unspecified described as a multicentre study **Prior testing:** lesions included in the study were selected because of their diagnostic difficulty does not specify what prior tests were done **Setting for prior testing:** Unspecified **Exclusion criteria:** Poor quality index test image (all images scoring 4 were excluded from the study) **Sample size (patients):** No. included: 40 patients **Sample size (lesions):** No. included: 43 **Participant characteristics:** Median age 39.5 years, (range 3–91 years). Male: 21 (53%); Female 19 (47%) **Lesion characteristics:** Site - Face 2; Head 1, Neck 1, Trunk 8, Arms 3, Legs 7, Back 20, Buttocks 1
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** **Method of diagnosis:** All lesions were examined with a dermatoscope during the face to face clinical diagnosis. Diagnosis was made by a expert dermatologist based on clinical features and dermoscopic findings. No specific algorithm (e.g. the Stolz index) was used for dermoscopic diagnosis. **Prior test data:** Unclear **Diagnostic threshold:** Not reported

	Diagnosis based on: Single (n=1) **Observer qualifications:** Dermatologist (an expert in the diagnosis of pigmented skin lesions) **Experience in practice:** High **Experience with index test:** High [Also evaluated Teledermatology assessment of transmitted images]

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Unclear
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard:** Histological diagnosis alone **Details:**All lesions were excised for a histopathological evaluation **Target condition (Final diagnoses)** - Melanoma (invasive): 11, BCC: 3 - Sebhorrheic keratosis: 2, Benign naevus: Melanocytic naevus 23, 'Benign' diagnoses: Angiokeratoma 1, lentigines 3
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	1. Excluded participants: Not reported 2. Time interval to reference test: Not reported 3. Time interval between index test(s): Not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Piccolo 2002

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR; 6-month period **Country** Italy
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented lesions excised because of equivocal dermoscopic findings or at the patient's request **Setting:** Secondary (general dermatology); from authors' institution **Prior testing:** Dermatoscopic suspicion; Patient request for evaluation/excision **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** No. included: 289 **Sample size (lesions):** No. included: 341 **Participant characteristics:** Mean age 33.6y, range 3–83y; Male gender: 127 (43.9%); Fitzpatrick phototype I to II (31.4%); Type III (42%); Type IV-V (26.4%) **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Clinical photographs and dermoscopic images. Cases were clinically and dermoscopically evaluated on a high-resolution colour monitor, in a random sequence **Prior test data:** None; appears to be based on images only **Diagnostic threshold:** Correct diagnosis of melanoma **Diagnosis based on:** Single observer (n=2) **Observer qualifications:** Dermatologist; (Dermatology?) resident **Experience in practice:** High - dermatologist had 5 years of experience; Low - resident with minimal training in PSLs **Experience with dermoscopy:** High and Low (resident had 6 months of experience, comprising 8 h of specialized training on three consecutive days and 2h per week in the routine of dermoscopy)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: "All excised lesions were examined histopathologically by a dermatopathologist" Disease positive: 13; Disease negative: 328 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 13 Sebhorrheic keratosis: 3; Benign naevus: 316; Dermatofibromas 7; angiomas 2
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	Time interval to reference test: nr
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	

Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Piccolo 2014

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** September 2010 to October 2013 **Country:** Italy
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Dermoscopically atypical pigmented skin lesions selected from the archives of the Dermatology Department at the University of L'Aquila, Italy; described as "a panel of ... retrospectively selected PSLs" **Setting:** Secondary (general dermatology) **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** Location/site of lesion - acral sites and the face **Sample size (patients):** No. included: 165 **Sample size (lesions):** No. included: 165 **Participant characteristics:** Mean age 43.5 yrs (range 12 to 84 years); Male gender: 59.4% **Lesion characteristics:** lesion site - upper extremities 18 (11%); lower extremities 53 (31%); 62 (37.5%) on the back; 32 (19.4%) on the chest. Melanoma thickness 87.9% (29/33) <0.75mm; 11% (4/33) >1.5 mm
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** ABCD **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used **Diagnostic threshold:** Total dermoscopic score (TDS) >4.75 and >5.45. **Diagnosis based on:** Single observer (n=4) **Observer qualifications:** 3 dermatologists and 1 GP with different degrees of dermoscopic experience **Experience in practice:** Mixed **Experience with dermoscopy:** High (Observer 1 - dermatologist); Moderate (Observers 2 and 3 - dermatologists); Low (Observer 4 - GP; underwent dermoscopic training by studying an interactive atlas of dermoscopy between time periods T0 and T1). **Any other detail:** Experience was scored based on number of years specializing in dermoscopy; number of pigmented skin lesions assessed by dermoscopy on a daily basis; number of relevant workshops/ seminars attended; and the number of authored publications on dermoscopy.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias

Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described) **Target condition (Final diagnoses)** Melanoma (invasive): 23; Melanoma (in situ): 10 Benign naevus: 105 Clark nevi; 19 Spitz/Reed nevi; 5 blue nevi; 3 dermal nevi.	
~~Is the reference standard~~ likely to correctly classify the target condition?		Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?		Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias		
Flow and timing	**Excluded participants:** none reported **Time interval to reference test:** not described	
Was there an appropriate interval between index test and reference standard?		Unclear
Did all patients receive the same reference standard?		Yes
Were all patients included in the analysis?		Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?		
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?		
Could the patient flow have introduced bias?		Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Pizzichetta 2002

Patient Selection

A. Risk of Bias		
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** April 1996 -Sept 1998 **Country** Italy	
Was a consecutive or random sample of patients enrolled?		Yes
Was a case-control design avoided?		Yes
Did the study avoid inappropriate exclusions?		No
Could the selection of patients have introduced bias?		High risk

B. Concerns regarding applicability		
Patient characteristics and setting	**Inclusion criteria:** Small (</=5mm) melanocytic skin lesions with "dermoscopic appearance not excluding melanona" that were surgically excised at the Centro di Riferimento Oncologico (National Cancer Institute), Aviano. **Setting:** Specialist unit; National Cancer Institute **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** size >5mm **Sample size (patients):** No. included: 123 **Sample size (lesions):** No. included: 129 **Participant characteristics:** Median age 30y, range 13 to 65y. Lesion site: trunk: 67 (52%); upper limbs/shoulder: 16 (14%); lower limbs/hip: 21 (16.3%); abdomen 21 (16.3%); foot 4 (3.1%) **Lesion characteristics:** median diameter 4mm (range: 1.2 to 5mm)	
Are the included patients and chosen study setting appropriate?		No
Did the study avoid including participants with multiple lesions?		Yes
Are there concerns that the included patients and setting do not match the review question?		High

Index Test

Index tests	**Dermoscopy** Pattern analysis; ABCD **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used; Only images assessed for presence/absence of dermoscopic criteria and dermoscopic diagnosis **Diagnostic threshold:** ABCD >5.45 and >=4.75; Pattern analysis: "Dermoscopic criteria used for evaluation were pigment network alterations, irregular extensions, branched streaks. gray·blue areas. pseudopods. brown globules. black dots, whitish blue veil, hypopigmentation, white scar-like areas and linear and dotted vascular patterns." **Diagnosis based on:** Single observer (n= 2) **Observer qualifications:** Not reported; likely oncologist/dermatologist (one observer based in Dept Oncology, other in Dermatology dept) **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: Histopathologic diagnosis of all specimens was performed by a single pathologist at the Department of Pathology of the Centro di Riferimento Oncologico Disease positive: 5 lesions Disease negative: 124 lesions **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 5 lesions Benign naevus: 124 lesions
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Each lesion imaged 'before surgery'
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Pizzichetta 2004

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** Jan 1996 to Dec 2001 **Country** Participants recruited from 5 participating centres (4 in Italy and 1 in USA) study conducted in Italy
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Clinical and/or dermoscopic hypomelanotic (extent of pigmentation <=30%) and amelanotic skin lesions seen and excised at the five participating centres **Setting:** Secondary (general dermatology) **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Not reported **Exclusion criteria:** Poor quality index test image (considered under Flow and TIming) **Sample size (patients):** No. included: 151 **Sample size (lesions):** No. eligible: 174; No. included: 151

	Participant characteristics: mean age 47 years (± 17.5 SD); male gender: 73 (48%) **Lesion characteristics:** Lesion site - head/neck (5.3%); trunk (20.5%); upper limbs/shoulder (11.9%); lower limbs/hip (25.2%); back (21.2%); abdomen (11.3%); hand (3.3%); foot (1.3%). Melanoma thickness: ≤1mm 85.3% (n=29); >1mm 14.7% (n=15)

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** Clinical photographs **Prior test data:** Only gender, age at diagnosis and the site of the skin lesion were known to the observer **Other test data**: File contained clinical and dermoscopic images; unclear whether both observed at the same time. **Diagnostic threshold:** investigated clinical features such as elevation, ulceration, shape, borders, colour **Diagnosis based on:** Single observer (n=1) **Observer qualifications:** Not reported; assumedDermatologist **Experience in practice:** Not described **Experience with index test:** Not described # **Dermoscopy** Pattern analysis **Method of diagnosis:** Dermoscopic images **Prior test data:** Clinical image also available **Diagnostic threshold: A**ssessed the lesions using the following dermoscopic criteria associated with melanoma and non-melanocytic skin lesions: pigment network, pigmentation, streaks, dots/globules, blue-whitish veil, regression structures, hypopigmentation, leaf-like areas, multiple grey-bluish globules, central white patch and vascular pattern **Test observers** as described for Visual Inspection (above) **Other detail:** Any other detail 122 images were taken with a digital stereomicroscope and 52 were taken with a Dermaphot camera (Heine Optotechnik; Herrsching, Germany) (· 10 magnification)and then digitalized with the Kodak PhotoCDsystem. Ultrasound gel was used on all the lesions (52)photographed with the Dermaphot in the Avianocentre. The other centres used the digital stereomicroscope consisting of a stereomicroscope and a Sony 3CCDDXC-930P colour video camera. The digital images were taken at a magnification of · 10–20.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone **Target condition (Final diagnoses)** Melanoma (invasive): 34 (39 in full sample); Melanoma (in situ): 5; Other diagnoses reported only for full sample of 151 (only 108 with clinical images for VI evaluation): 55 (40 with clinical images) "amelanotic/hypomelanotic non melanocytic lesions" (25 BCC, 4 SCC, 10 dermatofibroma, 8 Bowen's disease, 8 seborrhoeic keratosis) 52 (29 with clinical images) "amelanotic/hypomelanotic benign melanocytic lesions" (24 compound naevi, 17 dermal naevi, 5 Spitz naevi, 4 congenital naevi and 2 combined naevi).
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes

Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** 23 lesions excluded due to image quality; further 43 lesions were not available for evaluation by clinical images ("mainly benign melanocytic lesions"). **Time interval to reference test:** Not reported

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	**Time interval between index test(s):** not reported

Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Unclear
Are there any concerns that the test comparison could have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Pupelli 2013

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective **Period of data collection** 2007-2011 **Country** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Consecutively excised melanomas <5 mm diameter and 3 randomly sampled histologically proven small-diameter naevi for each included melanoma **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) [from Author institution] **Prior testing:** Selected for excision (no further detail); All had undergone dermoscopy and RCM in order to be included **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** lesion size >5mm excluded; Disagreement between evaluators on tumour histological classification **Sample size (patients):** No. included: 96 **Sample size (lesions):** No. included: 96 **Participant characteristics:** Mean age: melanoma group 48y (IQR 17, 77y); naevi 41y (IQR 6, 82y). Male gender: 54% of melanoma group; 58% of naevi group **Lesion characteristics:** Lesion site - trunk: 62% naevi; lower limbs/hip: 46% melanomas; Melanoma thickness: mean 0.37mm (SD 0.44mm). Lesion size (invasive melanoma): 77% (n=10) < 1 mm, 13% (n=3) >=1 mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy;** 7-point checklist **Method of diagnosis:** Dermoscopic images **Prior test data:** body site and age; it appears that RCM images also available at time of image interpretation. "For each lesion a complete set of dermoscopic and confocal images (including the whole lesion) was available"; "Dermoscopic and confocal microscopic images were evaluated – in blind from histological diagnosis, but not from the body site or the age of the patient". **Diagnostic threshold:** Score >=3 **Diagnosis based on:** Unclear likely single (n=NR) **Observer qualifications:** Not reported; no description of observers was provided **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: histopathology performed by two independent board-certified pathologists; disagreements were reviewed by both pathologists to obtain a consensus diagnosis. Disease positive: 24; Disease negative: 72 **Target condition (Final diagnoses)** Melanoma (invasive): 13; Melanoma (in situ): 11 Benign naevus: 72 (29 junctional, 19 compound, intra-dermal, eight blue, four lentigo simplex and seven Spitz)
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Rao 1997

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** not reported **Country** USA
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients with atypical melanocytic lesions or suspected early malignant melanoma **Setting:** Private care **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Private care **Exclusion criteria:** lesions over 13mm in diameter were excluded as they could not fit entirely within the standardized photographs **Sample size (patients):** No. included: 63 **Sample size (lesions):** No. included: 72 **Participant characteristics:** None reported **Lesion characteristics:** Melaoma thickness - ≤1mm: 100% of MM (n=21)
Are the included patients and chosen study setting appropriate?	No

Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection** ABCD **Method of diagnosis:** Clinical photographs and dermoscopic images **Prior test data:** Dermoscopic images also presented to observer but unclear whether both viewed at the same time or not; "Each color transparency was independently analyzed" by observers. The 1) clinical, 2) "overall" dermoscopic, and 3) ABCD "scored dermoscopic diagnoses of either MM or AMN were recorded for each lesion by the same observers. No indication of blinding between images **Diagnostic threshold:** Clinical variables were defined as follows: Asymmetry (A): Both silhouette and colour distribution were considered. Border irregularity (B): This was judged by the unevenness of the perimeter. Color (C): Color variegation and number of colours were evaluated. Diameter (D): The largest in situ diameter in millimetres of each lesion was recorded **Diagnosis based on:** Single observer (n=4) **Observer qualifications:** Two experienced dermatologists, and two melanoma fellows **Experience in practice:** Mixed experience (low and high experience combined) **Experience with dermoscopy:** Not reported # **Dermoscopy** ABCD and no algorithm **Method of diagnosis:** Clinical photographs and dermoscopic images **Prior test data:** Clinical examination and/or case notes The 1) clinical, 2) "overall" dermoscopic, and 3) ABCD "scored dermoscopic diagnoses of either MM or AMN were recorded for each lesion by the same observers. No indication of blinding between images **Diagnostic threshold:** ABCD scored dermoscopic diagnosis [Lesions with a score of 4.75 or less were classified as benign, those with scores 4.76 to 5.45 as suspicious, and those with scores of more than 5.45 as melanomas. Each feature was given a score of"1". Thus, the score ranged from 1 to 5] Overall dermoscopic diagnosis - no threshold reported; the overall dermoscopic impression was recorded based on criteria in the recently published textbook by Stolz et al. (1994 Color Atlas). **Test observers** as described for Visual Inspection (above) **Any other detail** All photographs were taken with the dermophot standard lens-to-lesion distance, aperture, and flash. Fujichrome 50 color 35 mm-transparency film was used and all exposed film was processed in the same laboratory (Colorite, New York, NY, USA)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: Each of the 72 melanocytic neoplasms was histopathologically diagnosed as with AMN or an early MM by a dermapathologist with special expertise in melanocytic neoplasms. Each lesion was completed excised and step sectioned. Disease positive: 21 MMs; Disease negative: 51 AMN **Target condition (Final diagnoses)** Melanoma (invasive): 21 51 atypical melanocytic nevus
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: none reported **Time interval to reference test**: not reported

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	**Time interval between index test(s)** : not reported

Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Rigel 2012

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Unclear **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR **Country** US

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Unclear
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions that had been analysed as part of a prior study using a MSDSLA system (Monheit 2011); melanomas and other pigmented lesions presumably selected on a case-control type basis **Setting:** Unclear **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** No. included: NR **Sample size (lesions):** No. included: 24 **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images; Interactive melanoma session where dermatologists were first presented with clinical and dermoscopic images and asked to make a diagnosis; then presented with information from MelaFind **Prior test data:** Patient history and clinical images were presented along with dermoscopic images **Diagnostic threshold:** Clinical decision to excise or not **Diagnosis based on:** Average (n=179) **Observer qualifications:** Dermatologist; practicing dermatologists attending an educational conference **Experience in practice:** Assumed High (median duration of practice 11-15 years) **Experience with dermoscopy:** Not reported

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes

If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone *Histology (not further described);* Disease positive: 5; Disease negative: 19 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 5; 'Benign' diagnoses: 19
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Rosendahl 2011

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** 30-month period; dates NR **Country** Australia
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Consecutive series of pigmented lesions submitted for histology from the primary care skin cancer practice of one author. **Setting:** Primary/private; skin cancer practice of one author. **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Primary **Exclusion criteria:** Poor image quality (considered under Flow and Timing) **Sample size (patients):** No. included: 389 **Sample size (lesions):** No. eligible: 466 pigmented lesions out of 1959 lesions excised or biopsied; No. included: 463 **Participant characteristics:** Mean age: 57y (SD 17). Male gender: 67.4% **Lesion characteristics:** (53.1%) melanocytic. Lesion site: 17.7% head or face; Trunk: 52.1%; 27.6% extremities; 2.2% palms or soles. Melanoma thickness: ≤1mm: 1/29 melanoma (3.4%)
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** Clinical photographs overview and close up image presented **Prior test data:** No further information used **Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone. **Diagnostic threshold:** Clinical diagnosis/subjective impression. Observers gave a diagnosis with level of confidence (from 0 for definitely benign to 100 for definitely malignant) after viewing the clinical images. (NB used authors threshold for detection of any skin cancer which includes lesions clinically

	considered to be MM, BCC pigmented epithelial carcinoma including SCC, keratoacanthoma, actinic keratosis and Bowen's disease as test positive; review only considered histologically confirmed MM, BCC or invasive SCC to be disease positive) **Diagnosis based on:** Single observer (n=NR) **Observer qualifications:** Expert dermatologist (based on author communication). **Experience in practice:** Expert **Experience with dermoscopy:** Expert # **Dermoscopy** Pattern analysis; new algorithm - Chaos and clues **Method of diagnosis:** Clinical photographs (one overview and one close-up), followed by one dermoscopic image presented to a blinded observer on a computer screen **Prior test data:** Clinical image only; Diagnosis made based on clinical image before presentation of dermoscopic image **Diagnostic threshold:** Observers gave a diagnosis with level of confidence (from 0 for definitely benign to 100 for definitely malignant). Chaos and clues short algorithm - each assessed for evidence of ''chaos'' (asymmetry of colour or structure); if present then ''clues'' searched for. Chaos - asymmetry of structure and colour defined according to the basic principles of pattern analysis as revised by Kittler (2007). Clues included: eccentric structure-less zone (any colour except skin colour), grey or blue structures, peripheral black dots or clods, segmental radial lines or pseudopods, polymorphous vessels, white lines,thick reticular or branched lines, and parallel lines on ridges (acral lesions). **Observers** as for visual inspection

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	No
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: Excise or biopsy Disease positive: 138; Disease negative: 325 **Target condition (Final diagnoses)** Melanoma (invasive): 9; Melanoma (in situ): 20; BCC: 72; cSCC: 5 (including 2 keratoacanthoma) 'Benign' diagnoses: 18 Bowen's disease and 14 actinic keratosis, 217 benign melanocytic plus additional 140 benign non melanocytic *authors considered Bowen's disease, actinic keratosis and keratoacanthoma as malignant; all considered benign for review analysis
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: Lesions were excluded due to poor image quality (n=3) **Time interval to reference test: U**nclear; lesions 'routinely photographed' if scheduled for excision or biopsy but not further described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	Time interval between index test(s): consecutive
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	High risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Rubegni 2012

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** Jan 2008- Dec 2010 **Country** Italy
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** All palmoplantar pigmented skin lesions observed and removed because of the presence of clinical and/or dermoscopic suspicious features and in the absence of any clear benignity pattern (parallel furrow pattern, lattice-like pattern or fibrillar pattern). **Setting:** Secondary (general dermatology) **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Secondary (general dermatology) **Exclusion criteria:** Non-acral lesions; site of lesion in volar skin of the folds near the toes; lesion size larger than 26mm diameter; non-melanocytic appearance; elevated or ulcerated appearance **Sample size (patients):** No. included: 107 **Sample size (lesions):** No. included: 107 **Participant characteristics:** Mean age: 49.8 years (women); 44.9 years (men); range 19 to 73 years; Male: 58.9%; Ethnicity white: 100% **Lesion characteristics** 78 on soles and 19 on palms; 9 (36%) melanomas <=0.75 mm (incl 4 in situ); 11 (44%) 0.76 to 1.5 mm in 11/25 lesions; 5 (20%) >=1.50mm
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Pattern analysis; 3-step algorithm for palmoplantar lesions (Koga 2011) **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used **Diagnostic threshold:** Clinical diagnosis (melanoma/no melanoma). For the 3-step algorithm the conventional options are "removal, follow-up or no follow-up"; the latter two were combined under the term 'no melanoma' for study purposes **Diagnosis based on:** Single observer (n=2; one per algorithm) **Observer qualifications:** Dermatologist **Experience in practice:** High **Experience with dermoscopy:** High; two dermatologists with 20 years' experience in dermoscopy **Any other detail:** ELM images achieved with the DB-Mips System; (magnification x 16),

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes

Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: "Histopathological diagnosis was based on the criteria of the National Institute of Health Consensus Conference" Disease positive: 25; Disease negative: 82 **Target condition (Final diagnoses)** Melanoma (invasive): 21; Melanoma (in situ): 4 'Benign' diagnoses: 82
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Rubegni 2016

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** 2010–2014 **Country** Not reported. Majority of authors based in Italy, but source of lesion images not described
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Consecutive melanocytic skin lesions showing clear-cut dermoscopic features of regression that were excised for suspected malignancy. Regression features included: blue-grey veil, blue grey globules and white scar-like areas, hypopigmented areas and atypical network (all of which may be present in benign and malignant lesions) **Setting:** Secondary; Not clearly reported but authors all based in Dermatology units or departments **Prior testing:** Dermatoscopic suspicion in all cases **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** No. included: 95 **Sample size (lesions):** No. included: 95 **Participant characteristics:** Median age: nevi group 36y (14 to 59 y); melanoma group 54.4y (17 to 89y). Male gender : 43; 45.2% **Lesion characteristics:** Lesion site: Head/Neck: 20 (40%) of Nevi; Trunk 23(46%) of nevi and 24 (55%) of melanoma group; extremities 7 (14%) of nevi group; other areas 20 (45%) of melanomas. Lesion size: mean 7.63mm, range 4 to 16mm (nevi) and 10.33mm 5 to 19mm (melanomas)
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Pattern analysis using 12 dermoscopic features of regression (study also developed a new classifier but data excluded from review due to use of leave one out procedure for validation) **Method of diagnosis:** Dermoscopic images; randomly presented to observers in blind to histopathological diagnosis. **Prior test data:** Unclear; data on morphology, site, age and gender were collected but not clear if presented along with image **Diagnostic threshold:** Diagnosis of melanoma or nevus following assessment of 12 dermoscopic structures suggestive of regression selected according to the literature (Zalaudek 2004; Seidenari 2010) including blue-grey areas, blue-whitish veil, blue globules and blue-grey peppering, white scar-like areas, white shiny streaks, atypical network, hypopigmented areas, irregular dots and globules, irregular streaks, irregular pigmented blotches and pink areas. **Diagnosis based on:** Single observer and consensus of 2/3 (n=3) **Observer qualifications:** Dermatologist

	Experience in practice: High; expert dermatologists **Experience with dermoscopy:** High; "experienced dermoscopists"

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone; Details: Every histological diagnosis was confirmed by 2 out of 3 expert dermopathologists Disease positive: 45; Disease negative: 50 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 45 Benign naevus: 50
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Sboner 2004

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR **Country** Italy (based on authors' institution)
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic lesion images acquired consecutively by d-ELM at the Department of Dermatology of Santa Chiara Hospital, Trento. **Setting:** Secondary (general dermatology)

	Prior testing: Selected for excision (no further detail) **Setting for prior testing:** Not reported **Exclusion criteria:** Seems that dysplastic nevi were excluded; "In this experimental setting, there were no dysplastic nevi" **Sample size (patients):** Not reported **Sample size (lesions):** No. included: 152 **Participant characteristics:** None reported **Lesion characteristics:** mean Breslow thickness for the invasive lesions is 1.0 +/- 0.7mm; 81% <=1.5mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images; digitial-ELM images presented on video device **Prior test data:** No further information used **Diagnostic threshold:** Not reported; appears to be correct diagnosis of melanoma **Diagnosis based on:** Single observer and average (n=8) **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Not described **Any other detail** The d-ELM Image Acquisition consists of a Leica WILD M-650 stereomicroscope (Leica Microsystem, Heerbrugg, Switzerland), with a SONY 3CCD DXC-930P colour camera (Sony Corporation,Tokyo, Japan). The software for image acquisition was DBDERMO MIPS (Dell'Eva/Burroni Studio, Florence/Siena, Italy). The digital image size has a spatial resolution of 768 x 576 pixels and a 24-bit colour resolution

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 42; Disease negative: 110 **Target condition (Final diagnoses)** Melanoma (invasive): 31; Melanoma (in situ): 11 Benign naevus: 110
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias

Comparative	

B. Concerns regarding applicability

Notes

Notes	

Seidenari 1998

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR; 4 year period **Country** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanomas and benign pigmented skin lesions from a larger series of pigmented skin lesions used to develop a new automated classifier; all melanomas with x20 magnification images were included plus a random sample of benign lesions with the same magnification. For the larger series, lesions were referred by dermatologists or general physicians because of one or more PSL that were difficult to interpret on clinical grounds alone, numerous PSLs, or because the patients were at increased risk for melanoma or had had a malignant PSL in the past. **Setting:** Secondary **Prior testing:** Clinical suspicion of malignancy **Setting for prior testing:** Primary; Secondary (general dermatology) **Exclusion criteria:** None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. eligible: 917; No. included: 100 **Participant characteristics:** None reported **Lesion characteristics:** Melanoma thickness: ≤1mm : 70.8% (n=46), <1 mm 58.5% (n=38). mean thickness 0.73 ± 0.69 mm;

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images; (obtained via videomicroscopy) **Prior test data:** No further information used; "Images appeared in a random sequence on the computer screen, and no information about the patient (such as history, skin site, age of the patient, evolution of the lesion) was given to the evaluators" **Diagnostic threshold:** Clinical diagnosis **Diagnosis based on:** Single observer (n=2) **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Low - one 'untrained' dermatologist; High - one routinely used videomicroscopy **Any other detail** For instrumental examination a 10- (39 cases), 20- (501 cases), or 50-fold–magnification (377 cases) was chosen according to the size of the lesion, enabling the whole lesion to be seen on the monitor. For the study, the 31 MM with x20 magnification were selected plus a random sample of 59 benign

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: describes using "conventional histopathologic criteria."

	Disease positive: 31; Disease negative: 59 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 31 'Benign' diagnoses: 59 "nonmelanoma cases consisted of nevi including dysplastic nevi"

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Seidenari 2005

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** Not reported **Country** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients with melanocytic lesions referred to a pigmented lesion clinic by a dermatologist for examination of a particular lesion or the whole skin; all lesions were excised for clinical, dermoscopic, or cosmetic reasons. **Setting:** Specialist unit **Prior testing:** Clinical and/or dermatoscopic suspicion; Patient request for evaluation/excision **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** NR **Sample size (lesions):** No. included: 603 **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** Pattern analysis **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used; images were retrospectively subdivided into four groups according to diagnoses performed exclusively by dermoscopy by two dermatologists trained in dermoscopy **Diagnostic threshold:** Images grouped according to degree of atypia, with those grade 3 considered to be melanomas dermoscopically, and those at grade 2 as dermoscopically atypical, to be excised to rule out melanoma **Diagnosis based on:** Not clear but appears to be consensus (2 observers) (n=2); "diagnoses performed exclusively by dermoscopy by two dermatologists trained in dermoscopy and experienced in using polarized light videomicroscopes" **Observer qualifications:** Dermatologist **Experience in practice:** Not reported **Experience with dermoscopy:** High Any other detail images were captured using a digital videomicroscope(VMS-110A, Scalar Mitsubishi, Tama-shi, Tokyo, Japan), with a 20-fold magnification.

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (no further details) Disease positive: 112; Disease negative: 491 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 112 Benign naevus: 491
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	Excluded participants: not reported Time interval to reference test: not reported Time interval between index test(s): not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Seidenari 2007

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR **Country** Italy **Test set derived** NR; The training set consisted of 369 melanocytic lesion images (including 43 MMs); test set comprised 243 images (including 43 MMs).
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Dermoscopic images of melanocytic lesion that had undergone excision **Setting:** Unclear **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Unspecified **Exclusion criteria:** None reported **Sample size (patients):** Not reported

	Sample size (lesions): No. eligible: 612; No. included: 243 in test set **Participant characteristics:** None reported **Lesion characteristics** MMs of the test set included 8 *in situ* with mean thickness was 0.77 mm.

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images observed on a computer screen **Prior test data:** No further information used; clinicians had no access to the clinical image or to clinical data **Diagnostic threshold:** Clinial diagnosis of melanoma **Diagnosis based on:** Single observer (n=4; results presented per observer, but not identifiable by experience level) **Observer qualifications:** Dermatology registrar (n=3); Dermatologist (n=1) **Experience in practice:** Not reported **Experience with dermoscopy:** Mixed: Trained (residents had undergone 6-month daily training on dermoscopy); High (dermatologist employed dermoscopy on a regular basis)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (no further details) Disease positive: 43; Disease negative: 200 **Target condition (Final diagnoses)** Melanoma (invasive): 35; Melanoma (in situ): 8 Benign naevus: 200
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Skvara 2005

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** July 1996 to Sept 1996 **Country** Austria

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Consecutive lesions excised due to changes over time during digital dermoscopy follow-up (appear to be from patients with multiple melanocytic naevi); all lesions were assessed for presence of dermoscopic characteristics and all melanomas plus random sample of same number of benign were assessed by dermoscopic algorithms (included in review) **Setting:** Secondary (general dermatology) **Prior testing:** Not reported **Setting for prior testing:** Unspecified **Exclusion criteria:** Location/site of lesion - palmar, plantar, facial lesions; lesion size lesions that exceeded maximum field of view of the electronic camera **Sample size (patients):** NR **Sample size (lesions):** No. included: 126 **Participant characteristics:** None reported **Lesion characteristics** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** 7-point checklist and ABCD **Method of diagnosis:** Dermoscopic images presented on a computer screen **Prior test data:** No further information used **Diagnostic threshold:** ABCD score >4.75; 7-point checklist score >2 **Diagnosis based on:** Single observer (n=2) **Observer qualifications:** Dermatologist **Experience in practice:** Assumed High; paper describes assessment of baseline images for dermoscopic criteria by "2 experienced dermatologists"; "additionally, the baseline images of (a subgroup of lesions) were evaluated by *2 blinded investigators*). These appear to be separate groups of observers but have assumed similar levels of experience. **Experience with dermoscopy:** Assumed High (as above)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (but all lesions followed up) Details: "standard histopathology" following lesion changes over time Disease positive: 63; Disease negative: 63 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 63 Benign naevus: 63

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Lesions suggestive of melanoma at baseline were removed at the patient's initial visit (immediately); the others were followed up for 3 to 6 months until lesion changes initiated excision
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Yes
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	Yes
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Soyer 1995

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Unclear **Period of data collection** Not reported **Country** Austria
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesion, difficult to diagnose on clinical grounds alone **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) **Prior testing:** Clinical suspicion **Setting for prior testing:** Secondary (general dermatology); referred by dermatologists or general physicians **Exclusion criteria:** None reported **Sample size (patients):** NR **Sample size (lesions):** No. included: 159 **Participant characteristics:** None reported **Lesion characteristics** "23 melanomas with a Breslow index of </= 0.75mm, 13 melanomas with a Breslow index >/=0.76mm and </= 1.5mm, 12 melanomas with a Breslow index >/=1.51mm and </=3.5mm, 2 melanomas with a Breslow index of >/=3.5mm."
Are the included patients and chosen study setting appropriate?	Unclear
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	Unclear

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Other test data**: Dermoscopy undertaken by same clinician(s) subsequent to clinical evaluation **Diagnostic threshold:** Not reported **Diagnosis based on:** n= 2 (1 or 2 per lesion) **Observer qualifications:** Dermatologist **Experience in practice:** Not clearly described; assumed to be High; "the examination was performed by a dermatologist expert in dermoscopy" **Experience with dermoscopy:** High; "the examination was performed by a dermatologist expert in dermoscopy" **Other detail:** "Photographic documentation was performed using an incident light stereomicroscope (Wild M 650) equipped with a Minolta XG-M camera" # **Dermoscopy** Pattern analysis **Method of diagnosis:** In person diagnosis "After application of a drop of immersion oil, each lesion was examined with a hand-held dermatoscope" **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** Criteria included: pigment network, irregular extensions, radial streaming, brown globules, black dots, whitish veil, white scar-like areas, gray-blue areas, hypopigmented areas, reticular depigmentation, amongst others. **Any other detail** "After application of a drop of immersion oil , each lesion was examined with a handheld dermatoscope (Heine, Optotechnik, Herrsching, Germany) at a magnification of x 10 and with an incident light stereomicroscope (Wild M 650, Heerburg, Switzerland) with 6- to 40-fold magnification."

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes

For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Unclear
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 65 (41%); Disease negative: 94 (59%) **Target condition (Final diagnoses)** Melanoma (invasive): 50; Melanoma (in situ): 15 BCC: pigmented basal cell carcinoma (3) Sebhorrheic keratosis: 18; Clark's nevus of dysplastic nevus (61 cases); lentigo actinica lentigo (2), pigmented actinic keratosis (4), angioma (3), angiokeratoma (2).
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: none reported **Time interval to reference test**: not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	tbc
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	Yes
Are there concerns that the test comparison differs from the review question?	Low concern

Notes

Notes	

Soyer 2004

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective (for expert observer data; previously acquired images prospectively interpreted by 6 inexperienced observers - data excluded as 3/6 medical students) **Period of data collection** Jan-Dec 2000 **Country** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Lesions at pigmented lesion clinic considered by experienced dermatologists to merit excision on clinical grounds. **Setting:** Specialist unit **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Specialist unit **Exclusion criteria:** None reported **Sample size (patients):** No. included: 225 **Sample size (lesions):** No. included: 231 **Participant characteristics:** Median age 34 years. Male gender: 110/225 (48.9%) **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm (study also develops 3-point checklist but data ineligible due to use of medical student observers) **Method of diagnosis:** In person **Prior test data:** Clinical examination **Diagnostic threshold:** Diagnosis of malignancy (melanoma or BCC) **Diagnosis based on:** Single observer (n= 1) **Observer qualifications:** Dermatologist **Experience in practice:** High; "experienced dermatologists" **Experience with dermoscopy:** High; "Each lesion was diagnosed dermoscopically by an experienced dermoscopist"

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Unclear
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described) Disease positive: 77; Disease negative: 154 **Target condition (Final diagnoses)** Melanoma (*in situ* and invasive, or not reported): 68 BCC: 9 'Benign' diagnoses: 154
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Appears consecutive; "before excision, each lesion was diagnosed dermoscopically"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes

If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Stanganelli 1998

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** Just states 1997 **Country** Italy

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Images of pigmented skin lesions selected from computerised files of the skin cancer clinic. **Setting:** Training study; images selected from skin cancer clinic **Prior testing:** Not reported **Setting for prior testing:** Unspecified **Exclusion criteria:** None reported **Sample size (patients):** **Sample size (lesions):** No. included: 30PSLs **Participant characteristics:** None reported **Lesion characteristics** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** Clinical photographs **Prior test data:** No further information used **Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone (images were randomised). **Diagnostic threshold:** Not reported **Diagnosis based on:** Average; n= 20 **Observer qualifications:** Dermatologist **Experience in practice:** Not described; **Experience with dermoscopy:** 30 dermatologists with “experience in ELM but (with) no formal training” attended a seminar on clinical and ELM diagnosis of PSL; 20 then participated in a test of their diagnostic accuracy. A second session on ELM was then held. # **Dermoscopy:** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** Post training, clinical image presented alongside dermoscopic image **Diagnostic threshold:** Not reported **Test observers** as described for Visual Inspection (above) # **Dermoscopy training:** Participants undertook 75 minute seminar on the overview of the principles of ELM using digital ELM (D-ELM) images from the files at the clinic. A second session 45 minutes long focused on the major aspects of the differential diagnosis of PSL as evaluated by D-ELM. **Length of training** 2 hrs **Post-training experience:** <6months **Training format** In-person teaching

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear

For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 10 BCC: 4 Mild//moderate dysplasia: 3; Sebhorrheic keratosis: 3; Benign naevus: Melanocytic nevi-7 Other: 1 hemangioma1 subungunal hemorraghe1 plantar intraepidermal hemorraghe
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: none reported **Time interval to reference test**: not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	**Time interval between index test(s):** not reported
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Unclear
Are there any concerns that the test comparison could have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Stanganelli 1999

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case-control (dermoscopy training study) **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** 15 Nov 1997-Jan 25 1998 **Country** Italy
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesion images (of melanomas, melanocytic naevi and non melanocytic naevi) selected from the dermoscopy files of two skin cancer clinics **Setting:** Specialist unit databases **Prior testing:** Not reported (all lesions excised) **Setting for prior testing:** Not reported

	Exclusion criteria: None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. included: 30 **Participant characteristics:** None reported **Lesion characteristics:** Melanoma thickness median 0.61mm, range 0.28-20mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm (training course covered principles of clinical and dermoscopic diagnosis of PSLs and referred to a number of diagnostic algorithms, however it did not teach any one particular method of diagnosis; same slides evaluated both pre- and post-dermoscopy training) **Method of diagnosis:** Clinical photographs and dermoscopic images **Prior test data:** No further information used; pairs of slides were projected onto a screen without access to patient information **Diagnostic threshold:** Correct diagnosis of melanoma **Diagnosis based on:** Average (n= 83 out of 465 professionals who participated in the meetings and workshops over the course of a year) **Observer qualifications:** Dermatologists **Experience in practice:** Mixed; "an average of 10y of general experience in dermatology (range 1-22yrs)" **Experience with dermoscopy:** Mixed; "A routine use of ELM was reported by 52 (63%) individuals". 35 (42%) see >20 PSLs per week # **Dermoscopy training:** Attendees could choose from several classes: Clinical classification and diagnosis of PSLs; Management of patients with PSLs; basic principles of ELM; ELM criteria; ELM diagnosis; Limitations of ELM. Length of training 4+2hr for each session attended **Training format** In-person teaching; delivered as one-day workshops and meetings

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Unclear
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described; original histological diagnosis used) Disease positive: 11; Disease negative: 19 **Target condition (Final diagnoses)** Melanoma (invasive): 10; Melanoma (in situ): 1 14 melanocytic nevi; 5 non-melanocytic lesions
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Stanganelli 2000

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective **Period of data collection** 1994-1996 **Country** Italy

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Patients with pigmented skin lesions referred by dermatologists and general practitioners either for pre-surgical assessment or consultation **Setting:** Specialist unit; "skin cancer clinic of Ravenna" **Prior testing:** patients referred for pre-surgical assessment or consultation indicating they have had prior tests **Setting for prior testing:** Primary some patients referred for consultation only; dermoscopy findings are reported back and management decision remains with referring clinician; Secondary (general dermatology) **Exclusion criteria:** None reported **Sample size (patients):** No. eligible: 1556 **Sample size (lesions):** No. eligible: 3372; No. included: 3372 **Participant characteristics:** Median age 30 years, range 10 to 94; Male: 522 (34%) **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	Yes
Did the study avoid including participants with multiple lesions?	No
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** ABCD **Method of diagnosis:** In person diagnosis **Prior test data:** N/A in-person diagnosis **Other test data**: Dermoscopic and clinical images subsequently presented separately to observer subsequent to diagnosis using clinical images alone. **Diagnostic threshold: NR** **Diagnosis based on:** Single observer; n= 1 **Observer qualifications:** Not reported; described as one of the co-authors and study based in skin cancer clinic - likely dermatologist **Experience in practice:** Not described **Experience with dermoscopy**: Not described **Other detail:** A crude clinical image (magn X6 and X10) was recorded in the digital database # **Dermoscopy:** Pattern analysis **Method of diagnosis:** Unclear; Patients seen in-person but dermoscopic diagnosis made based on digital ELM image (by same clinician as in-person clinical dx) **Prior test data:** Combined clinical/dermoscopy diagnosis **Diagnostic threshold:** Diagnosis described as based on an integrated synopsis of the patterns most commonly described in the literature (Steiner 1993) and generally associated with known histologic counterparts. Features were assessed described in detail with multiple references, including: presence of pigment network, sharp margins, abrupt edge of pigment network, branched streaks, pseudopods, radial streaming, brown globules, pigment dots, whitish or whitish blue veil, gray-blue areas, white or depigmented areas, maple leaf areas, milia-cysts, horny plugs and vascular patterns. **Test observers** as described for Visual Inspection (above) **Experience with dermoscopy:** Any other detail The equipment consisted of a Leica Wild M-650 stereomiscroscope (Leica AG, Heerbrugg, Switzerland), a Sony 3ccd DXC-930P color video camera, an AT-Vista videographics adapter, and IBM personal computer, a Sony Trinitron Analog PVM-2043MD monitor, and the DBDERMO MIPS software

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes

For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus follow up; histology report of known surgical excisions (n=262) plus a cancer-registry based follow up of benign cases (n=3110) **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 55; BCC: 43 'Benign' diagnoses: 3274
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: none reported **Time interval to reference test**: not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Yes
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	**Time interval between index test(s)**: not clearly reported just indicated that D-ELM was performed soon after clinical examination
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Unclear
Are there any concerns that the test comparison could have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	Unclear
Are there concerns that the test comparison differs from the review question?	Unclear

Notes

Notes	

Stanganelli 2005

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Unclear (likely case series) **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR **Country** Italy **Test set derived** A training set of 22 melanomas and 218 melanocytic nevi was randomised from the dataset. The test set was formed by the complement (the remaining 20 melanomas and 217 nevi). A further subset of images from the original dataset, consisting of 31 melanomas and 103 nevi, was used for the comparison between observers and CAD; derivation of the subset not reported.
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Unclear
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic lesions from patients referred to the Skin Cancer Unit and undergoing clinical and dermoscopic evaluation; images were 'selected' from a larger image database. Potential overlap with Stanganelli 2000 (not possible to determine). **Setting:** Specialist unit; Skin Cancer Unit in Ravenna **Prior testing:** Clinical and/or dermatoscopic suspicion **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** None reported

	Sample size (patients): No. eligible: 1556 referred / No. included: NR **Sample size (lesions):** No. eligible: 3274 / No. included: 477 melanocytic lesions; 237 in test set and 134 in comparison between CAD and human operators **Participant characteristics:** None reported **Lesion characteristics:** Melanoma thickness 61.2% <0.75mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** Clinical photographs **Prior test data:** General practitioners evaluated only clinical images; Dermatologists examined both clinical and dermoscopic images but unclear whether clinical diagnosis was made prior to presentation of dermoscopic images **Diagnostic threshold:** Not reported **Diagnosis based on:** Average (n=6) **Observer qualifications:** GP 3; Dermatologist 3 **Experience in practice:** NR **Experience with dermoscopy:** Assumed Low for GPs; High for dermatologists - described as “dermatologists with experience in ELM (2 years)” **Other detail:** Digital images included melanocytic lesions evaluated in ELM with a fixed x16 magnification # **Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images (dermatologists only) **Prior test data:** Dermatologists examined both clinical and dermoscopic images **Diagnostic threshold:** Not reported Test observers as described for Visual Inspection (above)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus cancer registry All included lesions underwent histology but some were identified using a cancer-registry-based follow-up of benign diagnoses. **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 42 in full sample; 31 in CAD vs human observer interp and 20 in test set 'Benign' diagnoses: 435 melanocytic nevi; 103 in CAD-observer comp and 217 in test set
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Unclear
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias

Flow and timing	**Excluded participants**: none reported **Time interval to reference test**: not reported

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Stanganelli 2015

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection/Prospective interpretation **Period of data collection** July 2010 to July 2012 **Country** Italy
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic lesions excised at the Skin Cancer Unit on the basis of clinical and/ or dermoscopic changes at follow-up suggesting a malignancy **Setting:** Specialist unit; "conducted at the Skin Cancer Unit at the 'Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori' (IRST IRCCS), in Ravenna/Forli and Meldola" **Prior testing:** Changes on digital monitoring; Lesions showing clinical or dermoscopic changes on follow-up **Setting for prior testing:** Specialist unit **Exclusion criteria:** Lack of baseline and follow-up dermoscopic images; lack of RCM images; lack of histology. **Sample size (patients):** No. included: 70 **Sample size (lesions):** No. included: 70 **Participant characteristics:** Mean age - women 39 years; men 40 years. Male gender: 54%. History of melanoma/skin cancer (37%). Total naevus counts, 27 (39%) with > 50 melanocytic naevi, 33 (47%) with 10–50 naevi; and 10 (14%) with <10 naevi. Fitzpatrick phototype I to II 19 (27%); Type III to IV 50 (73%). Median follow-up was 25 months (range 3–134 months) **Lesion characteristics:** Lesion site Head/Neck 7.1%; Trunk: 80%; Upper limbs/shoulder: 1.4%; Lower limbs/hip: 11.4%. Melanoma thickness median 0.4mm (0.2-1mm). Lesion size: mean at baseline 8 mm (range 2–22 mm); mean at FU 9 mm (range 3–24 mm)
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** Revised 7-point checklist (for follow-up purposes) **Method of diagnosis:** Dermoscopic images; baseline images assessed using standard 7-point checklist and compared to follow-up images to determine criteria indicating significant change **Prior test data:** Baseline dermoscopic image **Diagnostic threshold:** Presence of 'Major change' (asymmetrical structural and chromatic changes, or the appearance of melanoma-specific criteria, i.e. major or minor criteria on original seven-point checklist as per Argenziano 1998). Revised approach referenced to Argenziano 2010. ['Minor change' assigned if there was only symmetrical change in structural or chromatic pattern; 'moderate change' if either structural or chromatic changes were asymmetrical, but there were no melanoma-specific criteria; and 'no change' was assigned if all variables remained constant, with a tolerance of major axis change of 2 mm (Beer 2011; Terushkin 2012)] **Diagnosis based on:** Unclear; n= NR for dermoscopy **Observer qualifications:** Not reported but likely dermatologists (RCM images in same study were evaluated jointly by three expert dermatologists who had no knowledge of the clinical, dermoscopic or histopathology information) **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Details: "histopathological diagnosis was based on the consensus of at least two out of three board-certified pathologists" Disease positive: 12; Disease negative: 58 **Target condition (Final diagnoses)** Melanoma (invasive): 11; Melanoma (in situ): 1 'Benign' diagnoses: 55 melanocytic naevi (79%) and three nonmelanocytic lesions (4%).
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Appears consecutive; "Lesions showing clinical and/or dermoscopic aspects suggesting a malignancy are excised. RCM imaging is performed before surgical excision."
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Stolz 1994

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** From 1989 to 1991 **Country:** Germany **Test set derived:** 157 cases were randomly divided into a test and training set.
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Equivocal melanocytic skin lesions with size smaller than 9x13 mm, melanoma tumour thickness of <= 1mm and melanoma Clark's level III or less **Setting:** Secondary (general dermatology); Univerisity of Munich Department of Dermatology **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. eligible: 650 cases / No. included: 157 lesions **Participant characteristics:** None reported **Lesion characteristics:** Melanoma thickness: 50 <=0.4mm; 30 <=0.75mm; 15 <=1mm
Are the included patients and chosen study setting appropriate?	No

Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** ABCD **Method of diagnosis:** Dermoscopic images; colour prints examined for 31 dermoscopic features, most listed in the guidelines of the Consensus Conference of Surface Microscopy held in Hamburg in 1989 (Bahmer 1990); described as a "blind study" **Prior test data:** No further information used **Diagnostic threshold:** > 5.45; Multivariate analysis of training set data identified 8 features with the lowest p-values; the total dermoscopic score (TDS) was then developed based on: Asymmetry score x 1.3 + Border score x 0.1 +Colour score x 0.5 + Differential structure score x 0.5. New formula then evaluated on the test set of images. **Diagnosis based on:** Single observer (n=1) **Observer qualifications:** Not reported; co-author, assumed to be dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** - Histological diagnosis alone Details: Histology undertaken by two independent histopathologists Disease positive: Test set= 48; Disease negative: test set= 31 **Target condition (Final diagnoses)** Melanoma (invasive): 85; Melanoma (in situ): 10 'Benign' diagnoses: 62 melanocytic naevi; 17 junctional; 40 compound; 5 dermal
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Tan 2009

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control (dermoscopy training study) **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** NR **Country:** UK
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Test series of images of melanomas and benign lesions; source of images not reported **Setting:** Not described; Training images **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Unspecified **Exclusion criteria:** None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. included: 30 **Participant characteristics:** None reported **Lesion characteristics:** None reported
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** (Modified) pattern analysis **Method of diagnosis:** Clinical photographs and dermoscopic images **Prior test data:** Particpants presented with a test card printed on A4 laminated paper for each lesion, each consisting of one macroscopic and one dermatoscopic image **Diagnostic threshold:** Excise or not (algorithm not further described) **Diagnosis based on:** Average (n=6; all based at same University Hospital); the study authors presented 2x2 based on adding each 2x2 cell together for all observers; to avoid double counting of lesions for this review, all 2x2 cells were divided by 6 to get average result. **Observer qualifications:** Dermatology specialist registrar 3; Dermatologist 3 **Experience in practice:** Mixed **Experience with dermoscopy:** Low; Before the study, none had routinely used a dermatoscope. # **Dermoscopy training:** Participants received an online tutorial (http://www.dermatoscopy.org) teaching the Modified Pattern Analysis Diagnostic Algorithm (Steiner 1987; Carli 2003) and was given a dual polarizing LED dermatoscope to use in clinical practice for 10 months. At the end of the study, the test-card questionnaire was repeated. **Length of training** NR; online tutorial **Post-training experience:** 10 months **Training format** Online

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	No
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (not further described) Disease positive: 15; Disease negative: 15 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 15 Other: 15 (9 naevi, 1 blue naevus, 3 seborrhoeic keratoses, 1 lentigo and 1 vascular lesion).
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Tenenhaus 2010

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection:** Not described **Country** France
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Dermoscopic images of all melanoma lesions recorded on two pigmented skin lesion databases, plus random sample of benign naevus **Setting:** Secondary (general dermatology) **Prior testing:** Not reported **Setting for prior testing:** Unspecified **Exclusion criteria:** None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. included: 227 **Participant characteristics:** None reported **Lesion characteristics:** None reported
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** Clinical photographs **Diagnostic threshold:** Clinical diagnosis of melanoma and Excise decision; presence of ABCD and "malignancy-predictive" dermoscopic features were assessed (dichotomic answer) and diagnosis (melanoma, dysplastic or benign lesion) and therapeutic decision (dichotomic answer, excision/non-excision) given. **Diagnosis based on:** Single and Average (n=5); observers assessed lesion images independently; sensitivity and specificity also presented for 'pooled' advice **Observer qualifications:** Dermatologist **Experience in practice:** High; "senior dermatologists" **Experience with dermoscopy:** Not described; assumed High

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Unclear
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Unclear
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus other *Histology:* Excision and histopathology of lesions considered to be melanomas (n=32), dysplastic lesions (n=118) and some of those considered benign (n=15). Disease positive: 32; Disease negative: 165 Other: "lesions considered benign were not surgically excised"; assume observer diagnosis was used Disease positive: 0; Disease negative: 62 **Target condition (Final diagnoses)** Melanoma (invasive): 28; Lentigo maligna 4 Dysplastic nevus - 118; Blue benign naevus - 2;Congenitcal benign nevus - 5; junctional and dermic benign nevus - 7; Palmar-plantar benign nevus - 1; 'Benign naevus' not excised: 62
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	No
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	High

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Not described
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Unclear
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Unclear
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Troyanova 2003

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** Not reported **Country** Not reported
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Images of pigmented skin lesions selected for a dermoscopy training study **Setting:** Training study **Prior testing:** Not reported **Setting for prior testing:** Not reported **Exclusion criteria:** lesions that were >13mm were not included **Sample size (patients):** **Sample size (lesions):** No. included: 50 lesions **Participant characteristics:** None reported **Lesion characteristics:** Melanoma thickness: ≤1mm: 100%
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear

Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI):** No algorithm **Method of diagnosis:** Clinical photographs and dermoscopic images **Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone. **Prior test data:** No further information used **Diagnostic threshold:** Not reported **Diagnosis based on:** Average; n= 32 **Observer qualifications:** Dermatologist **Experience in practice:** High experience or 'Expert' **Experience with index test:** Low experience / novice users; experienced in PSL field but not ELM # **Dermoscopy:** No algorithm; possibly Pattern analysis **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used. Previously made diagnosis based on clinical images only; dermoscopic images presented after all clinical diagnoses had been made **Diagnostic threshold:** Not reported Test observers as described for Visual Inspection (above) **Dermoscopy training:** The group of 32 volunteer dermatologists had no formal training in the use of ELM, but had good theoretical knowledge and personal experience; participated in a teaching course comprised 6h of teaching on two consecutive days. The training was based on the presentation of several hundred slides with oral explanation of the ELM criteria. Tests were performed at the beginning and end of the teaching course. **Length of training** 2 days (12 hours in total) **Post-training experience:** <6months **Training format:** In-person teaching

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 25; Disease negative: 25 **Target condition (Final diagnoses)** Melanoma (invasive): 25 'Benign' diagnoses: 50 "not melanoma"
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: none reported

	Time interval to reference test: not reported

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	**Time interval between index test(s)**: not reported
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	Yes
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Unlu 2014

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** January 2008-January 2010 **Country** Turkey
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Melanocytic lesions excised at Ankara University Department of Dermatology Pigmented Lesion Clinic **Setting:** Specialist unit; Ankara University Department of Dermatology Pigmented Lesion Clinic **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** Location/site of lesion facial, nail and volar acral lesions were excluded; non-melanocytic appearance **Sample size (patients):** No. included: 115 **Sample size (lesions):** No. included: 115 **Participant characteristics:** Mean age: 38.72y (+/- 18.46 y). Male gender: n=56 (49%). **Lesion characteristics:** Lesion site: 100% trunk and limbs. Melanoma thickness: 10 (41.7%) <0.75mm; 14 (58.3%) >=0.75mm
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Yes
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm; Appears to be original clinical diagnosis at time of lesion presentation **Method of diagnosis:** In person diagnosis Appears to be diagnosis on presentation **Prior test data:** N/A in-person diagnosis **Other test data**: Dermoscopic images presented to different observers **Diagnostic threshold:** Not reported **Diagnosis based on:** Unclear - for visual inspection appears to be single examiner at time of clinic diagnosis (n=NR); dermoscopic images "scored by three other experienced dermatoscopists" **Observer qualifications:** Not reported; assumed dermatologists; described as experienced dermatoscopists **Experience in practice:** Unclear for clinic diagnosis; dermatoscopists described as "experienced" **Experience with index test:** Described as "experienced" # **Dermoscopy** 3-point rule; 7-point checklist; ABCD; CASH algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** No further information used; Clinical image evaluation appears to be separate from dermoscopy interpretation **Diagnostic threshold:** ABCD score- >=5.45 highly suggestive for melanoma; 7-point score- >=3; 3-point score- 2 or 3 criteria present; CASH algorithm- >=8 **Observers** as described for VI

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No

Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 24; Disease negative: 91 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 24 'Benign' diagnoses: 91 melanocytic benign lesions; 37 (32.2%) dermal nevi; 15 (13%) clark's nevi; 14 (12.2%) compound nevi; 13 (11.3%) blue nevi; 6 (5.2%) spitz nevi; 4 (3.5%) congenital melanocytic nevi; 2 (1.7%) junctional nevi
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: none reported **Time interval to reference test**: not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	**Time interval between index test(s):** Appear to be consecutively applied
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	Yes
Was the interval between application of the index tests less than one month?	Unclear
Are there any concerns that the test comparison could have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Viglizzo 2004

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Not reported **Period of data collection** Not reported **Country** Italy
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented skin lesions examined at the Dermoscopy Service and undergoing excision; a modified version of Kenet's risk stratification approach for dermoscopy (Ascierto 1998) was used to select high and very high risk lesions for excision; medium and low risk lesions were excised based

	on cosmetic or functional reasons. (2x2 data have been extracted only for melanocytic subgroup). **Setting:** Specialist unit (skin cancer/pigmented lesions clinic) Dermoscopy service at a university department (Department of Endocrinologic and metabolic disease) **Prior testing:** Clinical suspicion of malignancy without dermatoscopic suspicion **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** None reported **Sample size (patients):** No. eligible: 349 patients; No. included: not reported **Sample size (lesions):** No. eligible: 520 lesions; No. included: 79 lesions excised included in the final analysis **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI).** No algorithm **Method of diagnosis:** In person diagnosis **Prior test data:** Unclear **Diagnostic threshold:** Not reported; correct diagnosis of melanoma **Diagnosis based on:** Single observer (n=NR; "All dermoscopic evaluations were performed by the same operators") **Observer qualifications:** Not reported; "each lesion was .. diagnosed clinically and dermoscopically" at the Dermoscopy service **Experience in practice:** Not described **Experience with dermoscopy:** Not described; assumed High as diagnosis at 'Dermoscopy service' # **Dermoscopy** No algorithm; appears to be based on pattern analysis **Method of diagnosis:** In person diagnosis **Prior test data:** Clinical examination and/or case notes **Diagnostic threshold:** Lesion classification based on typical dermoscopic features: lesions with a pigment network and any of the classical dermoscopic features specific for melanoma, i.e. pseudopods, radial streaming or blue grey veil, were classified as very high risk. Lesions with a pigment network and dermoscopic features that might suggest melanoma but often seen in atypical nevi were classified as high risk. **Test observers:** as described above

Visual Inspection - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - in-person

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	Yes
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Yes
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias

B. Concerns regarding applicability

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone **Target condition (Final diagnoses)** Melanoma (invasive): 12 Melanocytic lesion: 67
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes

Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: none reported **Time interval to reference test**: not reported

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	**Time interval between index tests**: Not clearly reported but assumed consecutive as both recorded at Dermoscopy Service

Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	Yes
Are there concerns that the test comparison differs from the review question?	Low concern

Notes

Notes	

Wells 2012

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** NR **Country** US

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Pigmented lesions (melanomas and benign pigmented lesions) selected from a repository of lesions amassed during an acquisition study conducted by MELA Sciences Inc for the US Food and Drug Administration **Setting:** Company database (MELA Sciences Inc) of lesion images **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Not reported **Exclusion criteria:** None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. included: 47 **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** Clinical images and detailed clinical history; observers "viewed the images and a detailed case history for each lesion but were unaware of the MelaFind recommendations" **Diagnostic threshold:** Clinical diagnosis of melanoma or not; decision to biopsy the lesion **Diagnosis based on:** Average (n=39) **Observer qualifications:** Dermatologist **Experience in practice:** Not described **Experience with dermoscopy:** Not described

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone *Details*: "Lesions were biopsied in toto and evaluated by a panel of dermatopathologists who were unaware of the MelaFind recommendations" Disease positive: 23 / Disease negative: 24 **Target condition (Final diagnoses)** Melanoma (in situ and invasive, or not reported): 23 'Benign' diagnoses: 24
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Yes
Are there concerns that the target condition as defined by the reference standard does not match the question?	Low concern

Flow and Timing

A. Risk of Bias	
Flow and timing	**Participant exclusions:** None reported **Index test to reference standard interval:** Consecutive; "prior to biopsy of the lesion, photographs of the lesion were taken"
Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Low risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Westerhoff 2000

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control (for lesion selection; study was an RCT of dermoscopy training for PCPs) **Data collection:** Retrospective **Period of data collection** Not reported **Country:** Australia
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Clinically atypical pigmented skin lesions; 50 invasive melanomas and 50 non-melanomas randomly selected from the Sydney Melanoma Unit pigmented skin lesions (PSL) image database. **Setting:** Specialist unit (lesion selection) **Prior testing:** Selected for excision or followed up **Setting for prior testing:** Specialist unit (skin cancer/pigmented lesions clinic) **Exclusion criteria:** None reported **Sample size (patients):** No. included: NR **Sample size (lesions):** No. included: 100 **Participant characteristics:** None reported **Lesion characteristics:** median Breslow thickness 0.6mm
Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI):** No algorithm **Method of diagnosis:** Clinical photographs **Prior test data:** Unclear; all participants "were instructed not to look at the surface microscopic image until they had scored the clinical image" **Diagnostic threshold:** Not reported **Diagnosis based on:** Average (n=37; 74 practising primary care practitioners randomised to dermoscopy education intervention or not). **Diagnoses were recorded for both groups of GPs at baseline (pre-test) and after the training intervention had been administered to the intervention group (post-test), resulting in 8 sets of 2x2 data based on interpretation of the same set of 100 lesions; post-test data for the intervention group of GPs was used for the Visual Inspection analysis. **Observer qualifications:** GP **Experience in practice:** Considered to be Low; Only practitioners who had had no formal training with surface microscopy and did not use a surface microscope in their clinical practice were included. **Experience with dermoscopy:** Low experience / novice users (non-training arm); 'Trained' for the intervention arm **Other detail:** Any other detail Camera designed for close-up clinical photography (Elicar Macrolens, Japan) # **Dermoscopy:** No algorithm (non-training arm); Menzies criteria (training/intervention arm) **Method of diagnosis:** Dermoscopic images **Prior test data:** Diagnosis was first based on the clinical image and then the dermoscopic image for each lesion. **Diagnostic threshold:** Not reported; intervention arm instructed in Menzies criteria. **Test observers:** As above **Any other detail:** Dermoscopy at X10 magnification with a Dermphot camera (Heine Ltd) using oil at the skin-lens interface. # **Dermoscopy training:** The education intervention included provision of the Menzies et al pictorial atlas which reportedly describes the Menzies approach to dermoscopy diagnosis of melanoma (Menzies 1996); they also attended a 1-h presentation on dermoscopy reviewing the Menzies approach and including a quiz based on images of 25 different pigmented skin lesions **Post-training experience:** <6months; the median interval between pretest and education intervention was 46 days (range 5-155). median interval from education intervention to post-test was 23 days (range 2-54). **Training format** In-person teaching; Written materials

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	No
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis plus follow up *Histology:* All the lesions except two had been excised after photography and subjected to histopathological examination. Disease positive: 50 / Disease negative: 48 *Clinical FU plus histology of suspicious lesions:* The two benign PSL that had not been excised were monitored over a longer period of time and had shown no morphological change. Length of FU: NR; Disease positive: 0 / Disease negative: 2 **Target condition (Final diagnoses)** Melanoma (invasive): 50 / 'Benign' diagnoses: 50
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants:** none reported **Time interval to reference test:** "All the lesions except two had been excised after photography"

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Unclear
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	**Time interval between index test(s)**: Not reported; lesions described as "excised after photography" therefore assumed consecutive

Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	Yes
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Winkelmann 2016

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case control **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** not reported **Country** Not reported

Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	No
Did the study avoid inappropriate exclusions?	Unclear
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	**Inclusion criteria:** Images of pigmented skin lesions previously analysed by a digital classifier MSDSLA; method of selection of the 12 not reported **Setting:** Unclear; images selected for a Dermoscopy conference **Prior testing:** Not reported **Setting for prior testing:** Unspecified **Exclusion criteria:** None reported **Sample size (patients):** Not reported **Sample size (lesions):** No. included: 12 **Participant characteristics:** None reported **Lesion characteristics:** None reported

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Visual inspection (VI)** No algorithm **Method of diagnosis:** Clinical photographs **Prior test data:** Unclear **Other test data**: Dermoscopic images presented to observer subsequent to diagnosis using clinical images alone. **Diagnostic threshold:** Not reported - biopsy decision **Diagnosis based on:** Average (n=70) **Observer qualifications:** Dermatologist **Experience in practice:** Not described; recruited “dermatologists at a dermoscopy conference”. No further details **Other detail:** Authors report that practitioners with a particular interest in skin cancer or technology may have chosen to attend this conference and/or self-selected to take part in the study. # **Dermoscopy** No algorithm **Method of diagnosis:** Dermoscopic images **Prior test data:** Clinical images provided **Diagnostic threshold:** Not reported - biopsy decision **Test observers** as described for Visual Inspection (above)

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability	

Visual inspection - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	No
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone Disease positive: 5 / Disease negative: 7 **Target condition (Final diagnoses)** Melanoma (invasive): 3 / Melanoma (in situ): 2 Mild/moderate dysplasia: 7 low grade dysplastic nevi
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	
Flow and timing	**Excluded participants**: none reported **Time interval to reference test**: not reported
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	Yes
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	Unclear risk

Comparative

A. Risk of Bias	
Comparative	tbc
Was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	No
Was the interval between application of the index tests less than one month?	Yes
Are there any concerns that the test comparison could have introduced bias?	Low risk

B. Concerns regarding applicability	
Were all tests applied and interpreted in a clinically applicable manner?	No
Are there concerns that the test comparison differs from the review question?	High

Notes

Notes	

Zalaudek 2006

Patient Selection

A. Risk of Bias	
Patient Sampling	**Study design:** Case series **Data collection:** Retrospective image selection / Prospective interpretation **Period of data collection** February 2003 to January 2004 **Country** Naples, Italy
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low risk

B. Concerns regarding applicability	

Patient characteristics and setting	**Inclusion criteria:** Excised, equivocal and nonequivocal, pigmented and nonpigmented skin lesions with good image quality and melanin or haemoglobin pigmentation in all or part of the lesion. **Setting:** Specialist unit; specialized Pigmented Lesion Clinic database **Prior testing:** Selected for excision (no further detail) **Setting for prior testing:** Specialist unit **Exclusion criteria:** None reported **Sample size (patients):** NR **Sample size (lesions):** Eligible: 2621; Included - 150 (plus 15 lesions used for training purposes) **Participant characteristics:** None reported **Lesion characteristics** 37/165 (26%) considered equivocal on clinical and dermoscopic grounds **Thickness/depth:** Mean breslow 0.9mm

Are the included patients and chosen study setting appropriate?	No
Did the study avoid including participants with multiple lesions?	Unclear
Are there concerns that the included patients and setting do not match the review question?	High

Index Test

Index tests	**Dermoscopy:** 3 point checklist **Method of diagnosis:** Dermoscopic images, 'optimized for colour, brightness and contrast by using Adobe photoshop standards' **Prior test data:** Age, site, and gender provided **Diagnostic threshold:** >= 1 criterion present indicates malignancy (asymmetry - in colour and/or structure, not in shape; atypical network - pigment network with thick lines and irregular holes; and blue white structures - presence of any blue and/or white colour within the lesion **Diagnosis based on:** Average (n=150 out of 170 participating observers, who finished all 15 training cases and performed at least one evaluation of the main set of images (test set). Participation was open to all individuals regardless of professional profile and experience in dermoscopy; study was advertised through personal communication, e-mail correspondences, adverts during congresses and courses, as well as via the website (http://www.dermoscopy.org).). **Observer qualifications:** For full sample of 170: Dermatologists (n=125); GPs (n=15); Other professionals in the field of skin lesions (n=12); Medical students (n=7); Other medical specialty (n=11) **Experience in practice:** Not described **Experience with dermoscopy:** Mixed; 146/170 (86%) reported some experience with dermoscopy; 24 with no dermoscopy experience, 45 (26%) with >5 years experience. # **Dermoscopy training:** A web-based tutorial was provided to describe the concept of the three point checklist of dermoscopy including complete definitions of criteria and example images. Following web-based tutorial, observers initially scored a random sample of 15 images, receiving real-time feedback for that case as judged by an expert observer. **Training format:** Online

Visual Inspection - in-person

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - in-person

A. Risk of Bias

B. Concerns regarding applicability

Visual inspection - image-based

A. Risk of Bias

B. Concerns regarding applicability

Dermoscopy - image-based

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
For studies reporting the accuracy of multiple diagnostic thresholds, was each threshold or algorithm interpreted without knowledge of the results of the others?	
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Was the test applied and interpreted in a clinically applicable manner?	No
Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication?	Yes
Was the test interpretation carried out by an experienced examiner?	Unclear
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	**Reference standard** Histological diagnosis alone (no further details) **Target condition (Final diagnoses)** Melanoma (invasive): 18; Melanoma (in situ): 11 BCC: 18 79 melanocytic naevi; 26 seborrhoeic keratoses; 8 vascular tumours and 3 dermatofibromas
Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Expert opinion (with no histological confirmation) was not used as a reference standard	Yes
Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	Unclear
Are there concerns that the target condition as defined by the reference standard does not match the question?	Unclear

Flow and Timing

A. Risk of Bias	

Flow and timing	**Participant exclusions:** Poor quality index test image as exclusion criterion **Index test to reference standard interval:** Not described

Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
If the reference standard includes clinical follow-up of borderline/benign appearing lesions, was there a minimum follow-up following application of index test(s) of at least: 3 months for melanoma or cSCC or 6 months for BCC?	
If more than one algorithm was evaluated for the same test, was the interval between application of the different algorithms 1 month or less?	
Could the patient flow have introduced bias?	High risk

Comparative

A. Risk of Bias	
Comparative	

B. Concerns regarding applicability

Notes

Notes	

Footnotes

NR – not reported; PSL – pigmented skin lesion; PLC – pigmented lesion clinic; MM – malignant melanoma; MiS – melanoma *in situ* (or lentigo maligna); BCC – basal cell carcinoma; cSCC – cutaneous squamous cell carcinoma; LS – lentigo simplex; SK – seborrhoeic keratosis; SN – Spitz nevi; AK – actinic keratosis; BN – benign naevi; BD – Bowen's disease; DF – dermatofibroma; FU – follow-up; R –retrospective; P – prospective; CS – case series; CCS – case control study; WPC – within person comparison (of tests); BPC – between person comparison (of tests); NC – non comparative; RCM – reflectance confocal microscopy; CAD – computer-assisted diagnosis; FU - follow-up; ABCD(E) - asymmetry, border, colour, differential structures (enlargement); 7FFM - seven features for melanoma; 7PCL - seven point checklist; rev - revised; 3PCL - three point checklist; CASH - colour, architecture, symmetry and homogeneity

Characteristics of excluded studies

Ahnlide 2013

Reason for exclusion	EXCLUDE on index test *'clinical diagnosis' study*

Akasu 1996

Reason for exclusion	EXCLUDE on 2x2 data *no 2x2 data only describing the dermoscopic features present in the lesions*

Al Jalbout 2013

Reason for exclusion	EXCLUDE on sample size *case study*

Alendar 2009

Reason for exclusion	EXCLUDE on reference standard *only 7 reported verified histologically*

Altamura 2006

Reason for exclusion	EXCLUDE if individual lesion characteristics EXCLUDE on 2x2 data *looking for chars associated with acral melanoma; does not give 2x2 for overall dx*

Altamura 2010

Reason for exclusion	EXCLUDE on target condition (include for keratinocyte review only)

Amirnia 2016

Reason for exclusion	EXCLUDE on target condition (include for keratinocyte review only)

Antonio 2013

Reason for exclusion	EXCLUDE on target condition *Atypical nevi does not fall within our definition of D+*

Antoszewski 2015

Reason for exclusion	EXCLUDE on sample size *All excised lesions were benign.* EXCLUDE on 2x2 data

Aoyagi 2010

Reason for exclusion	EXCLUDE on sample size

Argenziano 1997

Reason for exclusion	EXCLUDE on study population *Only melanoma included*

Argenziano 1999

Reason for exclusion	EXCLUDE on study population *Only includes melanoma*

Argenziano 2002

Reason for exclusion	EXCLUDE not a primary study

Argenziano 2003

Reason for exclusion	EXCLUDE on 2x2 data *Table V gives se/sp data for 108 lesions but can't derive the number of melanoma for this subset of the original 128* EXCLUDE but contact authors; contacted 10-5-16 and 24-6-16

Argenziano 2004

Reason for exclusion	EXCLUDE if individual lesion characteristics *only lesions with vascular structures included; presence of 10 different characteristics assessed. 2x2 would be possible*

Argenziano 2004a

Reason for exclusion	EXCLUDE not a primary study *letter*

Argenziano 2008

Reason for exclusion	EXCLUDE on index test *surveillance/monitoring study*

Argenziano 2010

Reason for exclusion	EXCLUDE on index test *test used for follow-up looking at dermoscopic features of melanomas diagnosed 1 yr after follow up* EXCLUDE on 2x2 data

Argenziano 2011a

Reason for exclusion	EXCLUDE on target condition EXCLUDE on sample size *only 2 melanomas*

Argenziano 2011b

Reason for exclusion	EXCLUDE on target condition *5 melanoma metastases included as D+*

Argenziano 2012

Reason for exclusion	EXCLUDE on reference standard *no follow-up of test negatives*

Armstrong 2011

Reason for exclusion	EXCLUDE on reference standard *No reference standard results presented for the screened lesions; just compares naked eye judgements with dermoscopy*

Ascierto 1998

Reason for exclusion	EXCLUDE on 2x2 data *the data presented does not contribute to the review* EXCLUDE duplicate or related publication *Data included in Ascierto 2003*

Ascierto 2000

Reason for exclusion	EXCLUDE on 2x2 data EXCLUDE but contact authors *For excised lesions, study cross-tabulates ELM high/very high risk classification against some histological classification (Table 2). Number D+ = 580 (2x2: 504, 79, 76, 2072); 580 not mentioned anywhere else in paper [contacted* 10/05/2016 and 24/06/2016]

Ascierto 2003

Reason for exclusion	EXCLUDE not a primary study

Bafounta 2001

Reason for exclusion	EXCLUDE not a primary study *systematic review*

Bajaj 2016

Reason for exclusion	EXCLUDE on reference standard *unclear ref standard for benign diagnoses*

Bauer 2005

Reason for exclusion	EXCLUDE on index test *follow-up/monitoring study*

Bauer 2006

Reason for exclusion	EXCLUDE on index test *dermoscopy used to improve histopathology diagnosis*

Benati 2015

Reason for exclusion	EXCLUDE if individual lesion characteristics

Benelli 2000a

Reason for exclusion	EXCLUDE on 2x2 data *only inter-rater reliability data given (n=25); authors have published much larger evaluations of 7FFM and ABCD*

Benvenuto-Andrade 2006

Reason for exclusion	EXCLUDE on 2x2 data *diagnostic confidence rather than accuracy*

Benvenuto-Andrade 2007

Reason for exclusion	EXCLUDE on 2x2 data *agreement on lesion characterisation; not test accuracy*

Binder 1997

Reason for exclusion	EXCLUDE on 2x2 data *training study; only ROC curves/AUC presented pre and post-training* EXCLUDE but contact authors [contacted 10-5-16 and 24-6-16]

Blum 2003a

Reason for exclusion	EXCLUDE not a primary study

Blum 2004a

Reason for exclusion	EXCLUDE not a primary study *comment paper*

Blum 2004c

Reason for exclusion	EXCLUDE not a primary study *letter* *Letter only; limited data presented - evaluates '3-colour' rule as developed By Mackie 1992 (excluded as assessment of individual lesion features only)*

Blum 2004d

Reason for exclusion	EXCLUDE not a primary study *letter*

Blum 2006

Reason for exclusion	EXCLUDE on target condition *differentiates melanocytic from non-melanocytic lesions only*

Blum 2011

Reason for exclusion	EXCLUDE on study population *mucosal lesions only*

Blum 2014

Reason for exclusion	EXCLUDE on sample size *case studies*

Boespflug 2015

Reason for exclusion	EXCLUDE on study population *study aim is estimate the efficacy of an online spaced educational training for dermoscopy*

Bono 2001

Reason for exclusion	EXCLUDE on 2x2 data *aim of the study is to determine what features are present in amelanotic cutaneous melanoma*

Borsari 2010

Reason for exclusion	EXCLUDE if individual lesion characteristics EXCLUDE but contact authors *Paper focuses on diagnostic prediction of dermoscopic island for early melanoma, however the Methods describe the calculation of the total dermoscopy score and the 7-point checklist score; mean scores on each checklist per lesion type are then presented [no reply from authors]*

Bowns 2006

Reason for exclusion	EXCLUDE on index test; teledermatology study

Braun 2000

Reason for exclusion	EXCLUDE if derivation study *this is a pilot study on the new "wobble sign" in ELM no training/test sets used*

Braun 2007

Reason for exclusion	EXCLUDE if individual lesion characteristics

Braun-Falco 1990

Reason for exclusion	EXCLUDE on 2x2 data *Not a test accuracy study*

Brown 2000

Reason for exclusion	EXCLUDE not a primary study *systematic review*

Buhl 2012

Reason for exclusion	EXCLUDE on index test *follow up/monitoring* EXCLUDE duplicate or related publication *same patients as Haenssle 2010 #191*

Bystryn 2003

Reason for exclusion	EXCLUDE not a primary study *letter*

Cabrijan 2008

Reason for exclusion	EXCLUDE on 2x2 data *can't get 2x2; reports % correct diagnoses for each different lesion classification and not % misdiagnosed as melanoma or melanomas missed* EXCLUDE but contact authors *Study states "Dermatoscopic diagnosis were conformable with pathohistological diagnosis in 75 cases (72.82%) out of 103. The highest conformation was in diagnosing melanoma, in 5 out of 6 cases (83.3%)." which would give us sensitivity; do you have data on numbers mis classified as melanoma, i.e false positives? [author replied* 5-7-16 with some data but not sufficient to allow 2x2]

Canpolat 2011

Reason for exclusion	EXCLUDE if derivation study *looks at dermoscopic characteristics of acral lesions; only 4 suspicious lesions excised*

Cardenas 2009

Reason for exclusion	EXCLUDE on study population *Includes participants with palpable lesions; not all suspected of having skin cancer*

Carli 1998

Reason for exclusion	EXCLUDE on sample size *se/sp data are based on sample with only 4 MM*

Carli 2000

Reason for exclusion	EXCLUDE on target condition *only lesions histologically classified as common naevi or naevi with architectural disorder with/without cytological atypia were considered for the study.*

Carli 2003a

Reason for exclusion	EXCLUDE on sample size

Carli 2004

Reason for exclusion	EXCLUDE on sample size *<5 MM per arm* EXCLUDE on 2x2 data

Carli 2004a

Reason for exclusion	EXCLUDE on index test; can only estimate 2x2 for the full time period 1997 to 2001 across all observers, however dermoscopy was only introduced routinely in 1998 so some diagnoses prior to that will have been with visual inspection alone, and observers were classed as dermoscopy 'users' (those working in pigmented lesion clinics) and nonusers (general dermatology). EXCLUDE but contact authors *Author passed away; unable to make contact with co-authors*

Carli 2005

Reason for exclusion	EXCLUDE on 2x2 data EXCLUDE but contact authors *Study presents % MM correctly classified by naked eye +/- dermoscopy but doesn't give any detail on FPs, is thisavailable anywhere and/or are these lesions included in any subsequent publications? Author passed away; unable to make contact with co-authors*

Carlos-Ortega 2007

Reason for exclusion	EXCLUDE on 2x2 data *Gives se/sp for visual inspection and dermoscopy in the English abstract. 68 patients/70 lesions were included but only 36 seem to have had visual inspection results and all underwent dermoscopy. Two observers performed each test blinded to each other. Table I gives 22 with BCC and 11 with melanoma overall (no. D+ not reported for those with VI results), but using either or both of these numbers with the se/sp provided does not give the same PPV and NPV as given by the authors* EXCLUDE but contact authors *data not clearly presented for 2x2; translator suggested alternative but still does not work out to what is in paper; tried contacting authors twice, no reply as of* 28-07-16;

Carroll 1998

Reason for exclusion	EXCLUDE if derivation study *Derivation study; proposes new dermoscopic criteria for dx of BCC* EXCLUDE on 2x2 data

Chen 2013

Reason for exclusion	EXCLUDE on test observer

Ciudad-Blanco 2014

Reason for exclusion	EXCLUDE on study population *Includes melanoma only* EXCLUDE if individual lesion characteristics EXCLUDE on 2x2 data

de Giorgi 2006

Reason for exclusion	EXCLUDE on sample size *<5 cases of participants with a final melanoma diagnosis*

De Giorgi 2011

Reason for exclusion	EXCLUDE duplicate or related publication *Assesses same lesions as in Carli 2003b but different observers*

de Troya-Martin 2008

Reason for exclusion	EXCLUDE on study population *Only MM included*

Delfino 1997

Reason for exclusion	EXCLUDE if individual lesion characteristics EXCLUDE if derivation study EXCLUDE on 2x2 data *only reports association of each characteristics with D+/D-, not 2x2*

Di Chiacchio 2010

Reason for exclusion	EXCLUDE on target condition

	Excluding nail bed melanoma EXCLUDE on 2x2 data *There is insufficient data to extract for a 2x2 table*

Di Stefani 2007

Reason for exclusion	EXCLUDE on sample size *<5 malignant*

Dummer 1995

Reason for exclusion	EXCLUDE if individual lesion characteristics

Elwan 2016

Reason for exclusion	EXCLUDE on sample size EXCLUDE if derivation study EXCLUDE on 2x2 data

Fabbrocini 2008

Reason for exclusion	EXCLUDE on 2x2 data *there isn't sufficient data provided for each index test to populate 2x2 table* EXCLUDE but contact authors *As we can only include DTA studies - Do you have a cross tabulation of each clinician's diagnosis (e.g. at threshold of >=3 on 7 point checklist) against the histological diagnosis and/or a cross tabulation of the remote diagnosis against the Face to Face diagnoses? [author reply;* 30-6-16 cannot access data needed]

Ferrara 2002

Reason for exclusion	EXCLUDE on index test *this study looks at histopathological and dermoscopic disagreements not necessarily looking at how well dermoscopy differentiates between benign and malignant diagnosis*

Fidalgo 2003

Reason for exclusion	EXCLUDE on 2x2 data EXCLUDE duplicate or related publication *Appears to be superseded by Serrao 2006* EXCLUDE but contact authors *Paper provides % of MM and of DN with DNAOS scores of >=5.5 and >7, is it possible for you to provide the same information for the remaining 127 lesions in the study? Also can you advise as to whether any of the 247 lesions included in this study, overlap with the 652 reported in Serrao 2006 (#1144)? [author contacted 10-5-16; 24-06-16]*

Fruhauf 2012

Reason for exclusion	EXCLUDE on reference standard *35/219 underwent histology; 13 followed-up; 171 expert clinical Dx*

Fueyo-Casado 2009

Reason for exclusion	EXCLUDE on reference standard *<50% of the study population received histology as a test. No information given on those who were followed up.*

Giacomel 2005

Reason for exclusion	EXCLUDE on study population *Only BCC included*

Giacomel 2014

Reason for exclusion	EXCLUDE on sample size

Giannotti 2004

Reason for exclusion	EXCLUDE not a primary study *a review*

Gill 2015

Reason for exclusion	EXCLUDE on sample size EXCLUDE if derivation study

Gilmore 2009

Reason for exclusion	EXCLUDE if derivation study *Principle of lacunarity has been looked at before but not this particular application/approach to it* EXCLUDE on reference standard

	It is possible to get 2x2 for 'standard dermoscopy criteria' however dermoscopy negative were not excised and assumed benign; 201/312 underwent excision so theoretically eligible

Grichnik 2003

Reason for exclusion	EXCLUDE on sample size

Grichnik 2004

Reason for exclusion	EXCLUDE not a primary study *Editorial*

Guillod 1996

Reason for exclusion	EXCLUDE if derivation study

Gunduz 2003

Reason for exclusion	EXCLUDE on sample size *case study*

Hacioglu 2013

Reason for exclusion	EXCLUDE on target condition *Does not provide sufficient data for detection of melanoma*

Haenssle 2006

Reason for exclusion	EXCLUDE on index test *[surveillance study estimating accuracy of different approaches to follow-up]*

Haenssle 2010

Reason for exclusion	EXCLUDE on 2x2 data *Does not report specificity* EXCLUDE duplicate or related publication *same patients as Haenssle 2010 #191*

Haspeslagh 2016

Reason for exclusion	EXCLUDE if individual lesion characteristics EXCLUDE on 2x2 data

Henning 2007

Reason for exclusion	EXCLUDE if derivation study *First application of CASH algorithm*

Henning 2008

Reason for exclusion	Exclude is a derivation study

Herschorn 2012

Reason for exclusion	EXCLUDE not a primary study *systematic review*

Hirata 2011

Reason for exclusion	EXCLUDE on target condition EXCLUDE on index test

Hoffmann 2003

Reason for exclusion	EXCLUDE if derivation study *Uses leave one out cross validation procedure* EXCLUDE on 2x2 data *Only giving ROC values not able to extract a 2x2 table*

Hoorens 2016

Reason for exclusion	EXCLUDE on index test EXCLUDE on reference standard *No info on numbers undergoing histology; and no follow-up reported for benign appearing lesions* EXCLUDE on 2x2 data

Ishioka 2009

Reason for exclusion	EXCLUDE ON INDEX TEST - include for teledermatology only

Iyatomi 2006

Reason for exclusion	EXCLUDE if derivation study *uses leave one out procedure and same lesions and tumour extraction method as Iyatomi 2006* EXCLUDE on 2x2 data

Iyatomi 2008

Reason for exclusion	EXCLUDE if derivation study *the performance was evaluated by averaging both combinations (training and test sets) they did not present the data separately; uses leave one out procedure* EXCLUDE on 2x2 data *Not test accuracy; compares automated with manual extraction of tumour area*

Johr 2002

Reason for exclusion	EXCLUDE not a primary study

Kawabata 1998

Reason for exclusion	EXCLUDE if derivation study *aim of the study is to correlate findings between dermoscopy and histology findings of acral melanoma* EXCLUDE on 2x2 data *not test accuracy*

Kawabata 2001

Reason for exclusion	EXCLUDE on study population *MM of the nail bed*

Kefel 2012

Reason for exclusion	EXCLUDE if derivation study *no test set, first use of polarised light dermoscopy, various neural networks tested* EXCLUDE on 2x2 data

Kenet 1994

Reason for exclusion	EXCLUDE not a primary study EXCLUDE on 2x2 data *not an accuracy study*

Kittler 2002

Reason for exclusion	EXCLUDE not a primary study *systematic review*

Kittler 2006

Reason for exclusion	EXCLUDE conference abstract

Koga 2011

Reason for exclusion	EXCLUDE on reference standard *~23% of patients have their final diagnosis reached by histopathology 43/191*

Korotkov 2012

Reason for exclusion	EXCLUDE not a primary study *narrative review*

Lallas 2015

Reason for exclusion	EXCLUDE if derivation study *Develops new algorithm and does not use separate training/test sets of lesions*

Liebman 2011

Reason for exclusion	EXCLUDE not a primary study *comment*

Liebman 2012

Reason for exclusion	EXCLUDE not a primary study *comment*

Lipoff 2008

Reason for exclusion	EXCLUDE on target condition

	study does not differentiate MM from benign/other but looks to identify lesion characteristics that might help id those at risk for MM

Liu 2012

Reason for exclusion	EXCLUDE if derivation study *ásymmetry detection; 10-fold cross validation* EXCLUDE on 2x2 data

Lorentzen 1999a

Reason for exclusion	EXCLUDE on 2x2 data EXCLUDE but contact authors *Can you provide us with the number of melanomas that were included in the study so that we can estimate the 2x2 contingency tables using the se/sp data provided? Also is there overlap in the lesions included here with those included in the Lorentzen 2000 study? (see also author Qs for the 2000 study) [contacted 10-5-16; 24-6-16)*

Luttrell 2012

Reason for exclusion	EXCLUDE on test observer *Accuracy data only given for lay-persons not interested in this population of test observers*

MacKie 1971

Reason for exclusion	EXCLUDE on 2x2 data *only gives % with correct diagnosis rather than numbers misclassified as malignant*

MacKie 2002

Reason for exclusion	EXCLUDE if individual lesion characteristics *presence of 3 or more colours on dermoscopy*

Markowitz 2015

Reason for exclusion	EXCLUDE on target condition *Does not report sufficient data for detection of melanoma* INCLUDE based on full report

Massi 2001

Reason for exclusion	EXCLUDE if individual lesion characteristics

Mayer 1997

Reason for exclusion	EXCLUDE not a primary study *systematic review*

Menzies 1996a

Reason for exclusion	EXCLUDE if individual lesion characteristics *only given the SE/SP of individual characteristics; lesions make up the training set for Menzies 1996 (#1971)*

Menzies 1999

Reason for exclusion	EXCLUDE not a primary study

Menzies 2000

Reason for exclusion	EXCLUDE on target condition; BCC only

Menzies 2001

Reason for exclusion	EXCLUDE on index test *monitoring purposes*

Mun 2016

Reason for exclusion	EXCLUDE on reference standard *Only 37% of benign group underwent adequate reference standard*

Nathansohn 2007

Reason for exclusion	EXCLUDE on 2x2 data *Not test accuracy; follow-up study*

Navarrete-Dechent 2016

Reason for exclusion	EXCLUDE on target population; 2x2 for BCC only

Pan 2008

Reason for exclusion	EXCLUDE if derivation study *looking to id characteristics assoc with superficial BCC; 2x2 could be extracted for combination of 3 selected characteristics. Dermoscopic features selected based on prior studies but only patients with 3 diagnoses included: BCC, intra-ep carcinoma and psoriasis*

Panasiti 2009

Reason for exclusion	EXCLUDE if individual lesion characteristics EXCLUDE on reference standard *Of the 1543 lesions analysed on 321 received histopathology diagnosis. The accuracy data is based on this (only 20%) not sure what happened to the 80% of participants as no mention of follow up is mentioned.*

Pazzini 1996

Reason for exclusion	EXCLUDE on 2x2 data

Pehamberger 1987

Reason for exclusion	EXCLUDE on 2x2 data *Not test accuracy. This is a descriptive paper defining dermoscopic criteria. It is not a study testing accuracy of dermoscopy. From the authors final sign off it looks like part 2 of this paper may have details on accuracy (Steiner 1987).*

Pellacani 2002

Reason for exclusion	EXCLUDE not a primary study

Pellacani 2006

Reason for exclusion	EXCLUDE if derivation study *looks at detection of asymmetry between clinicians and computer* EXCLUDE on 2x2 data *2x2 could be derived for overall asymmetry or border cut-off but not overall diagnosis*

Pellacani 2007

Reason for exclusion	EXCLUDE if individual lesion characteristics EXCLUDE if derivation study *looking at blue hue*

Pellacani 2009

Reason for exclusion	EXCLUDE on target condition *focus is on identifying Spitz nevi from melanoma and 'clark' naevi and it is looking to derive useful RCM characteristics. Although some data is given in the text for an RCM score of >3 it is difficult to work out which are FP and which FN.*

Peris 2002

Reason for exclusion	EXCLUDE on study population *only patients with BCC diagnosis included*

Peris 2002a

Reason for exclusion	EXCLUDE not a primary study

Phan 2010

Reason for exclusion	EXCLUDE on 2x2 data *Not test accuracy investigating dermoscopic features of acral melanoma including of the nail apparatus; no accuracy data given*

Piccolo 2002a

Reason for exclusion	EXCLUDE not a primary study EXCLUDE on 2x2 data *not enough data to populate 2x2 table. No breakdown of index test results and ref standard.*

Piccolo 2004

Reason for exclusion	EXCLUDE on index test; *include for teledermatology anyway*

Piccolo 2006

Reason for exclusion	EXCLUDE on sample size *3 MMs, but also 1 lentigo and 14 dysplastic nevus; data not presented to allow se/sp estimation* EXCLUDE if individual lesion characteristics EXCLUDE if derivation study *Derivation for hypoluminescence microscopy;*

Pizzichetta 2001

Reason for exclusion	EXCLUDE on study population *population in study only those with malignant disease*

Pizzichetta 2001a

Reason for exclusion	EXCLUDE on 2x2 data *Observer agreement only*

Pizzichetta 2007

Reason for exclusion	EXCLUDE on study population *Only patients with melanoma included*

Pizzichetta 2010

Reason for exclusion	EXCLUDE on sample size *case study*

Pizzichetta 2013

Reason for exclusion	EXCLUDE if individual lesion characteristics *presence of negative pigmented network*

Pralong 2012

Reason for exclusion	EXCLUDE on study population *only melanoma pts included*

Provost 1998

Reason for exclusion	EXCLUDE on 2x2 data *Not test accuracy; only reports concordance*

Rader 2014

Reason for exclusion	EXCLUDE if individual lesion characteristics EXCLUDE on 2x2 data

Rajpara 2009

Reason for exclusion	EXCLUDE not a primary study *Systematic review*

Reggiani 2015

Reason for exclusion	EXCLUDE not a primary study systematic review kerationcyte skin cancer

Rigel 1997

Reason for exclusion	EXCLUDE not a primary study

Ronger 2002

Reason for exclusion	EXCLUDE if individual lesion characteristics

Rosendahl 2012

Reason for exclusion	EXCLUDE if individual lesion characteristics

Rosendahl 2012a

Reason for exclusion	EXCLUDE not a primary study

Rossi 2000

Reason for exclusion	EXCLUDE on reference standard *Unclear reference standard in disease negative*

Rubegni 2002

Reason for exclusion	EXCLUDE not a primary study

Rubegni 2005

Reason for exclusion	EXCLUDE not a primary study *Editorial*

Rubegni 2005a

Reason for exclusion	EXCLUDE not a primary study

Rubegni 2010

Reason for exclusion	EXCLUDE if derivation study *uses leave one out procedure* EXCLUDE on 2x2 data

Sahin 2004

Reason for exclusion	EXCLUDE if individual lesion characteristics EXCLUDE on 2x2 data *no accuracy data given, study looking at dermoscopic features of LM*

Saida 2002

Reason for exclusion	EXCLUDE if individual lesion characteristics *Decsriptive study looking at presence (%) of certain features. Not looking at accuracy. Has paragraph on diagnostic value of this specific feature quoting sens & spec but this is based upon unpublished observations and the data is not given in this paper.*

Saida 2004

Reason for exclusion	EXCLUDE if individual lesion characteristics

Sakakibara 2010

Reason for exclusion	EXCLUDE if individual lesion characteristics *only looking at different vascular structures*

Salerni 2011

Reason for exclusion	EXCLUDE on sample size *<5 cases*

Salerni 2012

Reason for exclusion	EXCLUDE on index test *surveillance study* EXCLUDE on 2x2 data

Salerni 2013

Reason for exclusion	EXCLUDE not a primary study *systematic review of surveillance with digital dermoscopy*

Salvio 2011

Reason for exclusion	EXCLUDE not a primary study EXCLUDE on sample size

Sanchez-Martin 2012

Reason for exclusion	EXCLUDE on study population *Only BCC cases*

Savk 2004

Reason for exclusion	EXCLUDE not a primary study *letter*

Sawada 2013

Reason for exclusion	EXCLUDE not a primary study

Sboner 2003

Reason for exclusion	EXCLUDE if derivation study *describes 10-fold cross-validation process for training/testing classifier*

Schulz 2001

Reason for exclusion	EXCLUDE on target condition *Melanoma metastases*

Scope 2015

Reason for exclusion	EXCLUDE not a primary study

Segura 2009

Reason for exclusion	EXCLUDE on index test; *RCM evaluation*

Seidenari 2004

Reason for exclusion	EXCLUDE on 2x2 data *No data to populate 2x2 table just ROC curve values given.* EXCLUDE but contact authors *TABLE 5 provides AUC values for each diagnosis for both formats and observers; we are particularly interested in accuracy for the diagnosis of melanoma, are you able to provide data in 2x2 format , e.g. for melanoma 'certain' against final diagnosis and for melanoma 'certain or fairly certain' against final diagnosis? [no reply from authors]*

Seidenari 2006

Reason for exclusion	EXCLUDE on study population *assessing best means of follow-in up patients with previous melanoma - total body exam versus only lesions >2cm. No melanoma identified*

Seidenari 2006a

Reason for exclusion	EXCLUDE if individual lesion characteristics *looks like this study is only looking at asymmetry judgement*

Seidenari 2012

Reason for exclusion	EXCLUDE if individual lesion characteristics *looks at indvdl lesion chars to distinguish Mel in situ, also gives mean ABCD and seven point scores* EXCLUDE on 2x2 data EXCLUDE but contact authors *Table 3 provides mean ABCD and seven point checklist scores, are you able to provide us with a cross tabulation of results with each checklist at 'standard' thresholds against final diagnosis? e.g. ABCD >4.75 and >5.45 for MIS and benoign groups 7-point checklist: presence >=2 chars and >=3 chars? [no reply]*

Seidenari 2013

Reason for exclusion	EXCLUDE on index test

Serrao 2006

Reason for exclusion	EXCLUDE on index test; *include for CAD review only*

Sgouros 2014

Reason for exclusion	EXCLUDE on index test; *include for CAD review only*

Shakya 2012

Reason for exclusion	EXCLUDE on target condition *SCC in situ is not included in target condition*

Shitara 2014

Reason for exclusion	EXCLUDE if individual lesion characteristics

Shitara 2015

Reason for exclusion	EXCLUDE on study population *includes only melanoma*

Sondak 2015

Reason for exclusion	EXCLUDE not a primary study *comment paper*

Soyer 1987

Reason for exclusion	EXCLUDE on 2x2 data *not test accuracy*

Soyer 2001

Reason for exclusion	EXCLUDE not a primary study *editorial*

Stanganelli 1998a

Reason for exclusion	EXCLUDE on 2x2 data *can't derive specificity; only gives 'exact diagnoses for MM and 2 benign categories and not number benign misdiagnosed as MM*

Steiner 1987

Reason for exclusion	EXCLUDE on 2x2 data; study only reports % correct diagnosis per lesion type for dermoscopy and does not list incorrect diagnoses

Steiner 1987a

Reason for exclusion	EXCLUDE on 2x2 data *only given the correct diagnosis for malignant* EXCLUDE but contact authors

Steiner 1993

Reason for exclusion	EXCLUDE if individual lesion characteristics EXCLUDE if derivation study

Stephens 2013

Reason for exclusion	EXCLUDE on sample size

Stoecker 2009

Reason for exclusion	EXCLUDE if derivation study *translucency* EXCLUDE on 2x2 data *data presented only as ROC curve and AUC*

Stoecker 2009a

Reason for exclusion	EXCLUDE not a primary study

Stoecker 2011

Reason for exclusion	EXCLUDE if individual lesion characteristics EXCLUDE if derivation study *Uses leave one out* EXCLUDE on 2x2 data *data presented only as ROC curve and AUC*

Stolz 2002

Reason for exclusion	EXCLUDE not a primary study

Stratigos 2007

Reason for exclusion	EXCLUDE on reference standard EXCLUDE on 2x2 data

Stricklin 2011

Reason for exclusion	EXCLUDE if individual lesion characteristics

Strumia 2003

Reason for exclusion	EXCLUDE conference abstract; letter only

Tasli 2012

Reason for exclusion	EXCLUDE not a primary study *systematic review looking at frequency of publications ion dermoscopy*

Teban 2003

Reason for exclusion	EXCLUDE on study population *classification of Clark nevi into 12 types* EXCLUDE on 2x2 data *No 2x2 data; classification of Clark nevi into 12 types*

Terstappen 2007

Reason for exclusion	EXCLUDE on study population *Includes only BCC - looking for BCC chars on Siascope* EXCLUDE if derivation study *Derivation study; first application of Siascope to pigmented BCC; 21/25 lesions were BCCs*

Terushkin 2010

Reason for exclusion	EXCLUDE on sample size *Only 2 invasive SCC*

	EXCLUDE on 2x2 data

Terushkin 2010a

Reason for exclusion	EXCLUDE on 2x2 data *Not test accuracy - reports final diagnoses of those excised over a number of time periods and benign-malignant ratio*

Tromme 2012

Reason for exclusion	EXCLUDE on reference standard *Inadequate ref test for disease negatives; expert dx only*

Tschandl 2012

Reason for exclusion	EXCLUDE on index test *Differentiating melanocytic from non-melanocytic lesions*

Tschandl 2015

Reason for exclusion	EXCLUDE on test observer *medical students*

Tschandl 2015a

Reason for exclusion	EXCLUDE if individual lesion characteristics

Ulrich 2015

Reason for exclusion	EXCLUDE on target condition *Does not provide sufficient data for evaluation of melanoma*

van der Leest 2011

Reason for exclusion	EXCLUDE on reference standard *Inadequate ref test for test negatives; expert dx only*

van der Rhee 2010

Reason for exclusion	EXCLUDE on reference standard *<50% of disease negative have an adequate reference standard*

van der Rhee 2011

Reason for exclusion	EXCLUDE on sample size *<5 cases*

Vasili 2010

Reason for exclusion	EXCLUDE conference abstract

Verduzco-Martinez 2013

Reason for exclusion	EXCLUDE on study population *Only BCC*

Vestergaard 2008

Reason for exclusion	EXCLUDE not a primary study *systematic review; check reference list*

Wang 2008

Reason for exclusion	EXCLUDE on 2x2 data *Not test accuracy; no details of misdiagnoses of benign lesions as malignant*

Warshaw 2009

Reason for exclusion	EXCLUDE on 2x2 data EXCLUDE duplicate or related publication Subgroup of participants from Warshaw 2010 EXCLUDE but contact authors *Study presents diagnostic accuracy of teledermatology and clinic diagnosis in comparison to histopathology; we need the underlying 2x2 contingency tables [see Warshaw 2010 for author response]*

Warshaw 2009a

Reason for exclusion	EXCLUDE on 2x2 data EXCLUDE duplicate or related publication Subgroup of participants from Warshaw 2010

	EXCLUDE but contact authors *Study presents diagnostic accuracy of teledermatology and clinic diagnosis in comparison to histopathology; we need the underlying 2x2 contingency tables [see Warshaw 2010 for author response]*

Warshaw 2010

Reason for exclusion	EXCLUDE on 2x2 data EXCLUDE but contact authors *Study presents diagnostic accuracy of teledermatology and clinic diagnosis in comparison to histopathology [author only able to provide numbers test positive and negative for melanoma and not for the final 2 cells of the 2x2; data provided showed higher sensitivity for melanoma as the primary diagnosis rather than as the 'aggregate' diagnosis and the 2x2 using the authors data and the accuracy figures from the paper showed more T+ from the primary diagnosis as opposed to the aggregate*

Weismann 2002

Reason for exclusion	EXCLUDE not a primary study

Wilkes 2010

Reason for exclusion	EXCLUDE not a primary study

Winkelmann 2015

Reason for exclusion	EXCLUDE duplicate or related publication

Winkelmann 2015a

Reason for exclusion	EXCLUDE duplicate or related publication

Witkowski 2016

Reason for exclusion	Exclude on target population; *include for keratinocyte review only*

Yadav 1993

Reason for exclusion	EXCLUDE on 2x2 data *Not test accuracy*

Yamaura 2005

Reason for exclusion	EXCLUDE if derivation study *gene amplification in acral lesions*

Yelamos 2016

Reason for exclusion	EXCLUDE not a primary study *commentary on Guitera 2016*

Yoo 2015

Reason for exclusion	EXCLUDE conference abstract

Youl 2007

Reason for exclusion	EXCLUDE on index test; evaluates 'clinical diagnosis' EXCLUDE but contact authors; author replied - dermoscopy used in some but not all lesions

Youl 2007a

Reason for exclusion	EXCLUDE on index test; evaluates 'clinical diagnosis' EXCLUDE but contact authors; author replied - dermoscopy used in some but not all lesions

Zaballos 2013

Reason for exclusion	EXCLUDE on study population *They do not have enough benign cases to include as full report.*

Zalaudek 2010

Reason for exclusion	EXCLUDE not a primary study *Editorial*

Zell 2008

Reason for exclusion	EXCLUDE on sample size *case study*

Zortea 2014

Reason for exclusion	EXCLUDE if derivation study

	Although data are divided into training and test sets, the test set data is used more than once over 20 realisations of each model, especially the melanomas, for which the same 10 are used in each realisation

Zou 2001

Reason for exclusion	EXCLUDE not a primary study *Study uses results from Stolz et al 1994* EXCLUDE on 2x2 data *Just showing ROC curves*

Footnotes

Characteristics of studies awaiting classification

Footnotes

Characteristics of ongoing studies

Footnotes

Summary of results tables

1 Summary of findings table

Question:	**What is the diagnostic accuracy of dermoscopy, in comparison to visual inspection, for the detection of cutaneous invasive melanoma and atypical intraepidermal melanocytic variants in adults?**		
Population:	Adults with lesions suspicious for melanoma, including: • those with limited prior testing (presenting in primary, community or private dermatology settings), and • referred populations (presenting in secondary care or specialist skin cancer clinics).		
Index test:	Dermoscopy with or without the use of any established algorithms or checklist to aid diagnosis, including: • in-person evaluations (face-to-face diagnosis), and • image-based evaluations (diagnosis based on assessment of a clinical image).		
Comparator test	Visual inspection		
Target condition:	Cutaneous invasive melanoma and atypical intraepidermal melanocytic variants		
Reference standard:	Histology with or without follow-up to confirm absence of malignancy in benign appearing lesions		
Action:	If accurate, positive results ensure melanoma lesions are not missed but are appropriately excised (or referred) and those with negative results can be safely reassured and discharged.		
	Number of studies	**Total lesions**	**Total melanomas**
Quantity of evidence	104	42,788	5700
Limitations			
Risk of bias: (in-person; image-based)	Potential risk for patient selection from use of case-control type design (19 image-based), inappropriate exclusion criteria (8; 25) or lack of detail (17; 27). All dermoscopy interpretation was blinded to reference standard diagnosis. Dermoscopy thresholds were clearly pre-specified (25; 50). Low risk for reference standard (29; 63); high risk from use of expert diagnosis or >20% of benign lesions with no histology (5; 11). Blinding of reference standard to clinical diagnosis reported only in one image-based evaluation. High risk for participant flow (15; 26) due to differential verification (6; 15), and exclusions following recruitment (10; 16). Timing of tests was not mentioned in 23 (18).		
Applicability of evidence to question: (in-person; image-based)	Participants restricted to those with melanocytic lesions only (10; 35) or other narrowly defined groups (5 image-based), or to those with histopathology results (29; 57) and included multiple lesions per participant (8 in-person). High concern for dermoscopy (16; 57) with no description of diagnostic thresholds (8; 25) or reporting of average or consensus diagnoses (9; 35). Dermoscopic image interpretation blinded to clinical images (51 image-based). Little information given concerning the expertise of the histopathologist (28; 50).		
FINDINGS:			
104 study publications (providing data for 103 cohorts of lesions) were included; 83 publications providing 86 datasets for evaluation of the primary target condition were separated a priori into in-person (n = 26) and image-based (n = 60) evaluations. Subsequent analysis confirmed differences in accuracy according to the different approaches to diagnosis (P < 0.0001). Analyses of studies by degree of prior testing revealed no obvious effect on accuracy; analyses were hampered by a lack of relevant information provided in the study publications and by the majority of studies apparently conducted in referred populations. The findings presented are based on results for all studies regardless of position on the clinical pathway. Sensitivities at fixed specificities and specificities at fixed sensitivities are given for *illustrative* purposes only and should not be taken as indicative of actual test performance.			
Test:	**In-person visual inspection alone versus visual inspection plus dermoscopy – any algorithm or threshold**		
Data analysed	Visual inspection	13 datasets - 6740 lesions; 459 cases	
	Dermoscopy	26 datasets - 23169 lesions; 1664 cases	

Results*	**Sensitivity (95% CI)**	**Fixed specificity**	**Fixed sensitivity**	**Specificity (95% CI)**
Visual inspection	76% (66, 85)	80%	80%	75% (57, 87)
Dermoscopy	92% (87, 95)			95% (90, 98)

Numbers applied to a hypothetical cohort of 1000 lesions**

	TP	**FN**	**FP**	**TN**	**TP**	**FN**	**FP**	**TN**
At a prevalence of 5%	VI: 38 **D: 46** ↑ 8	VI: 12 **D: 4** ↓ 8	190	760	40	10	VI: 238 **D: 47** ↓191	VI: 713 **D: 904** ↑191
At a prevalence of 12%	VI: 91 **D: 110** ↑19	VI: 29 **D: 10** ↓ 19	176	704	96	24	VI: 220 **D: 44** ↓176	VI: 660 **D: 836** ↑176
At a	VI: 160	VI: 50	158	632	168	42	VI: 198	VI: 5935

prevalence of 21%	**D: 193** ↑ 33	**D: 17** ↓ 33					**D: 40** ↓158	**D: 750** ↑158
Test:	**Image-based visual inspection alone versus visual inspection plus dermoscopy – any algorithm or threshold**							
Data analysed	Visual inspection				11 datasets - 1740 lesions; 305 cases			
	Dermoscopy				60 datasets - 13475 lesions; 2851 cases			
Results	**Sensitivity (95% CI)**		**Fixed specificity**		**Fixed sensitivity**		**Specificity (95% CI)**	
Visual inspection	47% (34, 59)		80%		80%		42% (28, 58)	
Dermoscopy	81% (76, 86)						82% (75, 87)	
Numbers applied to a hypothetical cohort of 1000 lesions* **								
	TP	**FN**	**FP**	**TN**	**TP**	**FN**	**FP**	**TN**
At a prevalence of 18%	VI: 85 **D: 146** ↑ 61	VI: 95 **D: 34** ↓ 61	164	656	144	36	VI: 476 **D: 148** ↓328	VI: 344 **D: 672** ↑328
At a prevalence of 24%	VI: 113 **D: 194** ↑81	VI: 127 **D: 46** ↓ 81	152	608	192	48	VI: 441 **D: 137** ↓304	VI: 319 **D: 623** ↑304
At a prevalence of 39%	VI: 183 **D: 316** ↑ 133	VI: 207 **D: 74** ↓ 133	122	488	312	78	VI: 354 **D: 110** ↓244	VI: 256 **D: 500** ↑244
Test:	**Results according to algorithm used to assist dermoscopy interpretation**							
	Datasets	**Lesions; cases**	**Sensitivity (95%CIs)**	**Specificity (95%CI)**	**Numbers in a cohort of 1000 lesions******			
					TP	**FN**	**FP**	**TN**
In-person					**At median prevalence of 12%**			
No algorithm	8	4704; 849	88% (75, 95)	87% (80, 92)	106	14	114	766
Pattern analysis	6	4307; 296	92% (87, 95)	92% (68, 98)	110	10	70	810
ABCD at > 5.45 (or likely)	5	1438; 160	81% (62, 92)	92% (82, 97)	97	235	70	810
Image-based					**At median prevalence of 24%**			
No algorithm	24	4498; 941	76% (70, 82)	79% (71, 85)	182	58	61	699
Pattern analysis	20	4621; 989	83% (76, 88)	87% (80, 92)	199	41	99	661
ABCD at > 5.45	7	2471; 406	81% (60, 92)	81% (69, 89)	194	46	144	616
7PCL at >= 3	11	3408; 798	80% (63, 91)	67% (51, 80)	192	48	251	509
3PCL	7	1505; 365	74% (61, 85)	60% (42, 76)	178	62	304	456

Footnotes

V - visual inspection; D - dermoscopy; CI - confidence interval; FU - follow-up; NR - not reported; ABCD(E) - asymmetry, border, colour, differential structures (enlargement); 7PCL - seven point checklist; 3PCL - three point checklist; TP - true positive; FP - false positive; FN - false negative; TN - true negative

* Numbers for a hypothetical cohort of 1000 lesions are presented for two illustrative examples of points on the SROC curves: firstly for the sensitivities of tests at fixed specificities of 80%; and secondly for the specificities of tests at fixed sensitivities of 80%.

**Numbers estimated at 25th, 50th (median) and 75% percentiles of invasive melanoma or atypical intraepidermal melanocytic variants prevalence observed across 26 datasets reporting in-person evaluations of dermoscopy added to visual inspection.

***Numbers estimated at 25th, 50th (median) and 75% percentiles of invasive melanoma or atypical intraepidermal melanocytic variants prevalence observed across 60 datasets reporting diagnosis using dermoscopic images

***Numbers estimated at median prevalence (50th percentile) of invasive melanoma or atypical intraepidermal melanocytic variants observed across 26 datasets reporting in-person evaluations of dermoscopy added to visual inspection and then for 60 datasets reporting diagnosis using dermoscopic images

Additional tables

1 Investigation of effect of pathway positions for detection of invasive melanoma or intraepidermal melanocytic variants

Test	Studies	Lesions (Cases)	DOR (95% CI)	Specificity at 80% sensitivity (95% CI)	Sensitivity at 80% specificity (95% CI)	Relative DOR (95% CI)	P-value (LR)
a. Pathway in-person							
2 Limited prior testing (all lesions included)							
2-c	2	566 (37)	15.2 (1.8, 128)	78% (20, 98)	79% (38, 96)	0.41 (0.03, 5.9)	0.001
4 Referred (all lesions included)							

4-c	2	3830 (82)	494 (58, 4218)	100% (94, 100)	98% (91, 100)	13.4 (1.06, 169)	
4-u	2	8764 (82)	111 (16.4, 765)	98% (80, 100)	95% (76, 99)	3.0 (0.25, 36.0)	
5 Referred (selected on reference standard)							
5-c	5	3247 (767)	36.9 (9.1, 150)	91% (64, 98)	88% (73, 96)	1.0 (Comparator)	
5-u	12	3847 (539)	77.0 (34.0, 174)	96% (90, 99)	93% (86, 97)	2.1 (0.41, 10.7)	
5* Referred (equivocal lesions only)							
5*-c	2	227 (70)	74.2 (6.4, 859)	96% (53, 100)	93% (66, 99)	2.0 (0.14, 29.3)	
7 Lesions undergoing follow-up							
7-u	1	2688 (87)	8.3 (0.63, 111)	63% (6, 98)	69% (21, 95)	0.23 (0.01, 4.7)	
b. Pathway (image based)							
3 Limited prior testing (selected on reference standard)							
3-c	1	45 (9)	0.39 (0.02, 8.2)	5% (0, 68)	14% (1, 69)	0.02 (0.001, 0.43)	0.007
3-u	1	463 (29)	7.5 (0.61, 92.8)	61% (8, 97)	67% (19, 95)	0.33 (0.02, 5.0)	
4 Referred (all lesions included)							
4-c	1	134 (31)	11.6 (0.94, 142)	73% (13, 98)	75% (25, 96)	0.51 (0.03, 7.7)	
4-u	4	1619 (248)	15.1 (4.2, 54.0)	78% (45, 94)	79% (55, 92)	0.66 (0.13, 3.5)	
5 Referred (selected on reference standard)							
5-c	6	1336 (304)	22.7 (8.0, 64.6)	85% (63, 95)	84% (68, 93)	1.0 (Comparator)	
5-u	35	7436 (1680)	16.0 (10.2, 25.0)	79% (70, 87)	80% (73, 85)	0.70 (0.23, 2.1)	
5* Referred (equivocal lesions only)							
5*-c	3	1210 (139)	84.0 (16.2, 436)	96% (79, 99)	94% (80, 99)	3.7 (0.52, 26.1)	
5*-u	6	956 (326)	49.4 (16.4, 149)	93% (79, 98)	91% (80, 96)	2.2 (0.47, 10.0)	
7 Lesions undergoing follow-up							
7-u	3	276 (85)	2.3 (0.50, 10.4)	29% (6, 72)	42% (16, 73)	0.10 (0.02, 0.63)	

Footnotes

DOR - diagnostic odds ratio; CI - confidence interval; c – clearly position on clinical pathway; u – unclear position on clinical pathway

1 Test naïve; 2 Limited prior testing; 3 Limited prior testing (with selection on reference standard); 3* Limited prior testing (with selection on reference standard and equivocal nature of lesions); 4 Referred for further assessment; 5 Referred for further assessment (with selection on reference standard); 5* Referred for further assessment (with selection on reference standard and equivocal nature); 6 Referred for further assessment (equivocal on specialist review); 7 Lesions that have been undergoing follow-up

2 Comparison of visual inspection and dermoscopy for detection of invasive melanoma or intraepidermal melanocytic variants

Test	Studies	Lesions (cases)	DOR (95% CI)	Specificity at 80% sensitivity (95% CI)	Sensitivity at 80% specificity (95% CI)	Relative DOR (95% CI)	P-value (LR)
In-person evaluations							
Visual inspection	13	6740 (459)	13.1 (7.0, 24.5)	75% (57, 87)	76% (66, 85)	4.7 (3.0, 7.5)	< 0.001
Visual inspection +Dermoscopy	26	23,169 (1664)	61.7 (34.9, 109)	95% (90, 98)	92% (87, 95)		
Change with adding dermoscopy to VI (95% CI)				+20% (+7, +33)	+16% (+8, +23)		
In-person evaluations (direct studies)							
Visual inspection	11	5854 (412)	13.7 (5.9, 31.8)	75% (49, 90)	77% (63, 87)	4.8 (2.8, 8.1)	< 0.001
Visual inspection +Dermoscopy	11	5854 (412)	65.7 (27.0, 160)	96% (87, 99)	92% (84, 96)		
Change with adding dermoscopy to VI (95% CI)				+21% (+2, +39)	+15% (+7, +23)		
Image based evaluations							

Clinical (macro) images	11	1740 (305)	3.2 (1.9, 5.4)	42% (28, 58)	47% (34, 59)	5.6 (3.7, 8.5)	< 0.001
Dermoscopic images	60	13,475 (2851)	17.8 (12.3, 25.7)	82% (75, 87)	81% (76, 86)		
Change replacing VI with dermoscopy (95% CI)				+40% (+27, +57)	+35% (+24, +46)		
Image based evaluations (direct studies)							
Clinical (macro) images	11	1740 (305)	3.6 (1.7, 7.6)	48% (25, 73)	47% (30, 64)	5.3 (3.5, 8.0)	< 0.001
Dermoscopic images	11	1735 (306)	19.2 (8.7,42.0)	83% (70, 91)	83% (68, 92)		
Change replacing VI with dermoscopy (95% CI)				+34% (+15, +53)	+36% (+20, +52)		

Footnotes

DOR - diagnostic odds ratio; CI - confidence interval; VI - visual inspection; LR - likelihood ratio test

3 Investigations of sources of heterogeneity in person studies positions for detection of invasive melanoma or intraepidermal melanocytic variants

Test	Studies	Lesions (cases)	DOR (95% CI)	Specificity at 80% sensitivity (95% CI)	Sensitivity at 80% specificity (95% CI)	Relative DOR (95% CI)	P-value (LR)
Difference in-person and image based							
In-person	26	23169 (1664)	73.2 (41.2, 130)	95% (92, 98)	94% (90, 97)	4.6 (2.4, 9.0)	< 0.001
Image	60	13475 (2851)	15.8 (10.7, 23.3)	79% (72, 86)	80% (73, 85)		
Difference (95% CI)				+16% (+9, +23)	+14% (+8, +21)		
Differences in reference standard (in-person studies)							
Histology	18	5105 (767)	51.4 (24.6, 107)	94% (86, 98)	91% (84, 95)	0.27 (0.06, 1.22)	0.23
Histology+FU	7	17733 (865)	188 (50.8, 697)	99% (93, 100)	97% (90, 99)		
Difference (95% CI)				+5% (-1, +10)	+6% (-0, +12)		
Use of an algorithm (in-person studies)							
No algorithm	16	9302 (1159)	72.6 (30.1, 175)	96% (88, 99)	93% (86, 97)	1.4 (0.34, 5.6)	0.17
Any algorithm	10	13867 (505)	52.3 (18.1, 151)	94% (82, 98)	91% (80, 96)		
Difference (95% CI)				-2% (-10, +7)	-2% (-11, +7)		
Lesion type (in-person studies)							
Melanocytic	8	2460 (416)	38.2 (11.7, 124)	91% (72, 98)	89% (74, 96)	0.48 (0.12, 2.0)	0.60
Pigmented	18	20709 (1248)	79.1 (35.7, 175)	96% (90, 99)	94% (88, 97)		
Difference (95% CI)				+5% (-7, +16)	+5% (-6, +15)		
Single or multiple individuals making diagnosis (in-person studies)							
Single	13	8436 (1044)	60.3 (21.7, 168)	95% (83, 99)	92% (83, 96)	1.0 (0.18, 5.8)	0.30
Consensus	7	12377 (294)	59.2 (14.9, 236)	95% (78, 99)	92% (76, 98)		
Difference (95% CI)				0% (-10, +10)	0% (-11, +11)		
Prevalence (in-person studies)							
0-5%	6	15392 (206)	99.1 (24.6, 400)	97% (87, 99)	94% (82, 98)	5.4 (0.80, 36.6)	0.008
>5-10%	6	1718 (117)	18.3 (4.7,71.9)	81% (43, 96)	81% (58, 93)	1.0 (comparator)	
>10-20%	6	2089 (312)	49.4 (14.0, 175)	94% (75, 99)	90% (77, 96)	2.7 (0.42, 17.3)	

>20%	8	3970 (1029)	92.1 (27.5, 309)	96% (86, 99)	93% (86, 97)	5.0 (0.78, 32.4)	

Footnotes

DOR - diagnostic odds ratio; CI - confidence interval; FU - follow-up; LR - likelihood ratio test

4 Investigations of sources of heterogeneity in image-based studies for detection of invasive melanoma or intraepidermal melanocytic variants

Test	Studies	Lesions (cases)	DOR (95% CI)	Specificity at 80% sensitivity (95% CI)	Sensitivity at 80% specificity (95% CI)	Relative DOR (95% CI)	P-value (LR)
Differences in reference standard (image based studies)							
Histology	48	10267 (2210)	20.8 (13.6 31.9)	84% (77, 89)	84% (77, 89)	2.8 (0.92, 8.9)	0.19
Histology+FU	8	2762 (549)	7.3 (2.6, 20.9)	64% (36, 84)	65% (41, 84)		
Difference (95% CI)				-20% (-47, +6)	-18% (-42, +5)		
Use of an algorithm (image based studies)							
No algorithm	42	8762 (1834)	18.9 (11.8, 30.3)	83% (74, 88)	82% (75, 88)	1.4 (0.60, 3.3)	0.22
Any algorithm	18	4713 (1017)	13.4 (6.7, 27.0)	77% (62, 87)	77% (63, 87)		
Difference (95% CI)				-6% (-20, +9)	-5% (-19, +9)		
Lesion type (image based studies)							
Melanocytic	30	6980 (1710)	18.1 (9.8, 33.4)	82% (70, 90)	82% (71, 89)	1.10 (0.49, 2.50)	0.16
Pigmented	26	4062 (733)	16.4 (9.6, 27.9)	80% (70, 88)	80% (71, 87)		
Difference (95% CI)				-2% (-15, 12)	-1% (-13, +10)		
Single or multiple individuals making diagnosis (image based studies)							
Single	26	5877 (1030)	27.2 (14.5, 51.2)	88% (78, 93)	87% (78, 92)	1.9 (0.80, 4.4)	0.30
Consensus	28	5232 (1350)	14.4 (8.1, 25.7)	78% (66, 87)	78% (68, 86)		
Difference (95% CI)				-10% (-23, +4)	-8% (-20, +3)		
Prevalence (image based studies)							
0-20%	14	4855 (519)	422 (65.2, 2726)	99% (94, 100)	98% (93, 100)	30.7 (1.51, 6.24)	0.12
>20-30%	17	3893 (901)	13.7 (1.2, 162)	78% (25, 97)	77% (20, 98)	1.0 (comparator)	
>30-40%	9	974 (330)	19.5 (8.8, 42.8)	83% (69, 91)	83% (68, 92)	1.4 (0.11, 18.8)	
>40%	14	1387 (630)	15.5 (0.2, 23.3)	79% (72, 85)	79% (71, 86)	1.1 (0.09, 13.9)	

Footnotes

DOR - diagnostic odds ratio; CI - confidence interval; FU - follow-up; LR - likelihood ratio test

5 Algorithm and threshold analysis for each definition of the target condition

Target condition Test	Datasets (n)	Lesions (Cases)	Pooled Sensitivity (95% CI)	Pooled Specificity (95% CI)	Datasets (n)	Lesions (Cases)	Pooled Sensitivity (95% CI)	Pooled Specificity (95% CI)
a. Invasive melanoma and atypical intraepidermal melanocytic variants	**IN-PERSON**				**IMAGE-BASED**			
no algorithm - any threshold	8	4707 (849)	0.88 (0.75, 0.95)	0.87 (0.80, 0.92)	24	4498 (941)	0.76 (0.70, 0.82)	0.79 (0.71, 0.85)
no algorithm - correct diagnosis	-	-	-	-	18	4118 (795)	0.77 (0.69, 0.83)	0.84 (0.76, 0.89)
no algorithm - excise decision	-	-	-	-	10	831 (263)	0.79 (0.69, 0.86)	0.55 (0.50, 0.61)
pattern – any threshold or NR	6	4307 (296)	0.92 (0.87, 0.95)	0.92 (0.68, 0.98)	20	4621 (989)	0.83 (0.76, 0.88)	0.87 (0.80, 0.92)

pattern - at >=1 char present	1	220 (33)	0.88 (0.72, 0.97)	0.79 (0.73, 0.85)	-	-	-	-
pattern - at >=3 chars present	1	68 (5)	1.00 (0.48, 1.00)	0.56 (0.42, 0.68)	-	-	-	-
pattern - correct dx	-	-	-	-	19	4095 (896)	0.81 (0.73, 0.87)	0.87 (0.80, 0.92)
pattern - excise decision	-	-	-	-	3	933 (227)	0.97 (0.68, 1.00)	0.72 (0.60, 0.81)
ABCD at NR (likely > 5.45)	1	235 (5)	1.00 (0.48, 1.00)	0.90 (0.85, 0.93)	-	-	-	-
ABCD at >5.45	4	1203 (155)	0.78 (0.58, 0.90)	0.93 (0.79, 0.98)	7	2471 (406)	0.81 (0.60, 0.92)	0.81 (0.69, 0.89)
ABCD at or likely > 5.45 (two previous groups combined)	5	1438 (160)	0.81 (0.62, 0.92)	0.92 (0.82, 0.97)	-	-	-	-
ABCD at > 4.75	1	309 (73)	0.83 (0.69, 0.92)	0.45 (0.39, 0.51)	10	4242 (816)	0.81 (0.67, 0.90)	0.72 (0.93, 0.80)
rev ABCD at >= 4	-	-	-	-	1	269 (84)	0.87 (0.78, 0.93)	0.89 (0.83, 0.93)
ABCD at 60% specificity	1	356 (73)	0.90 (0.81, 0.96)	0.60 (0.54, 0.66)	-	-	-	-
ABCD at 70% specificity	1	356 (73)	0.85 (0.75, 0.92)	0.70 (0.64, 0.75)	-	-	-	-
ABCD at 75% specificity	1	356 (73)	0.85 (0.75, 0.92)	0.75 (0.69, 0.80)	-	-	-	-
ABCD at 80% specificity	1	356 (73)	0.77 (0.65, 0.86)	0.80 (0.75, 0.84)	-	-	-	-
ABCD at 85% specificity	1	356 (73)	0.71 (0.59, 0.81)	0.85 (0.80, 0.89)	-	-	-	-
ABCD at 90% specificity	1	356 (73)	0.64 (0.52, 0.75)	0.90 (0.86, 0.93)	-	-	-	-
ABCDE at > 1.3	1	356 (73)	1.00 (0.95, 1.00)	0.15 (0.11, 0.20)	-	-	-	-
ABCDE at > 2.65	1	356 (73)	0.97 (0.90, 1.00)	0.39 (0.33, 0.45)	-	-	-	-
ABCDE at > 3.05	1	356 (73)	0.95 (0.87, 0.98)	0.57 (0.51, 0.62)	-	-	-	-
ABCDE at > 3.6	1	356 (73)	0.90 (0.81, 0.96)	0.70 (0.64, 0.75)	-	-	-	-
ABCDE at > 4.25	1	356 (73)	0.82 (0.71, 0.90)	0.82 (0.77, 0.86)	-	-	-	-
ABCDE at > 4.9	1	356 (73)	0.74 (0.62, 0.84)	0.90 (0.86, 0.93)	-	-	-	-
ABCDE at >= 4	-	-	-	-	1	269 (84)	0.90 (0.82, 0.96)	0.87 (0.81, 0.92)
7FFM at >= 2	1	401 (60)	0.80 (0.68, 0.89)	0.89 (0.85, 0.92)	4	2200 (340)	0.89 (0.76, 0.96)	0.84 (0.78, 0.89)
7PCL at >= 2	1	638 (108)	0.93 (0.86, 0.97)	0.98 (0.97, 0.99)	-	-	-	-
7PCL at >= 3	2	11137 (127)	0.67 (0.46, 0.83)	0.96 (0.88, 0.99)	11	3408 (798)	0.80 (0.63, 0.91)	0.67 (0.51, 0.80)
7PCL at >= 5	-	-	-	-	1	322 (70)	0.67 (0.55, 0.78)	0.83 (0.78, 0.87)
7PCL at NR	-	-	-	-	4	1936 (360)	0.72 (0.56, 0.84)	0.79 (0.61, 0.90)
rev 7PCL at NR (likely >= 1)	-	-	-	-	1	1678 (238)	0.61 (0.54, 0.67)	0.88 (0.86, 0.89)
rev 7PCL at >=1	-	-	-	-	1	300 (100)	0.88 (0.80, 0.94)	0.51 (0.44, 0.58)
rev 7PCL for FU - major change	-	-	-	-	1	70 (12)	0.67 (0.35, 0.90)	0.60 (0.47, 0.73)
Menzies at 2 neg and >= 1 pos	1	206 (23)	0.83 95)	0.69 (0.62, 0.75)	4	1856 (317)	0.78 (0.38, 0.96)	0.63 (0.39, 0.81)
Menzies at NR	-	-	-	-	2	60 (26)	0.77	0.82

							(0.57, 0.89)	(0.66, 0.92)
3PCL at >= 2	-	-	-	-	7	1505 (363)	0.74 (0.61, 0.85)	0.60 (0.42, 0.76)
4point (scored 3PCL) at > 2	-	-	-	-	1	75 (32)	0.84 (0.67, 0.95)	0.81 (0.67, 0.92)
Hofman algorithm at NR	-	-	-	-	1	254 (75)	0.87 (0.77, 0.93)	0.88 (0.82, 0.92)
CASH at >= 6	-	-	-	-	1	477 (119)	0.78 (0.70, 0.85)	0.51 (0.46, 0.56)
CASH at >= 8	-	-	-	-	2	190 (56)	0.97 (0.79, 1.00)	0.69 (0.60, 0.76)
Chaos/Clues at = 2	-	-	-	-	2	940 (148)	0.82 (0.75, 0.87)	0.53 (0.36, 0.70)
Acral 3-step	-	-	-	-	1	107 (25)	0.96 (0.80, 1.00)	0.91 (0.83, 0.96)
b. Invasive melanoma	**IN-PERSON**				**IMAGE-BASED**			
No algorithm - threshold NR	3	190 (62)	0.87 (0.76, 0.93)	0.96 (0.91, 0.98)	6	683 (202)	0.77 (0.59, 0.88)	0.79 (0.63, 0.90)
pattern analysis - threshold NR	1	45 (16)	0.81 (0.54,0.96)	0.97 (0.82, 1.00)	1	119 (24)	1.00 (0.86, 1.00)	0.97 (0.91, 0.99)
ABCD at > 4.2	1	495 (23)	0.88 (0.69, 0.97)	0.64 (0.60, 0.68)	-	-	-	-
ABCD at > 4.75	-	-	-	-	2	330 (85)	0.76 (0.66, 0.84)	0.84 (0.73, 0.91)
ABCD at > 5.45	2	832 (242)	0.79 (0.74, 0.84)	0.90 (0.58, 0.98)	1	258 (64)	0.45 (0.33, 0.58)	0.94 (0.89, 0.97)
7PCL at NR	-	-	-	-	1	332 (217)	0.90 (0.85, 0.94)	0.79 (0.71, 0.86)
Menzies at 2 neg and >= 1 pos	-	-	-	-	4	4184 (715)	0.91 (0.83, 0.96)	0.71 (0.68, 0.74)
3PCL at > NR	-	-	-	-	1	332 (217)	0.82 (0.77, 0.87)	0.40 (0.31, 0.50)
Kenet (modified) at MM likely	1	54 (12)	1.00 (0.74, 1.00)	0.95 (0.84, 0.99)	1	258 (64)	0.75 (0.63, 0.85)	0.94 (0.89, 0.97)
Kenet (modified) at MM possible	1	54 (12)	1.00 (0.74, 1.00)	0.45 (0.30, 0.61)	1	258 (64)	0.89 (0.79, 0.95)	0.87 (0.82, 0.91)
CASH at >= 8	-	-	-	-	1	332 (217)	0.82 (0.76, 0.86)	0.72 (0.63, 0.80)
Kreusch algorithm	-	-	-	-	1	265 (96)	0.98 (0.93, 1.00)	0.83 (0.77, 0.89)
Menzies for amelanotic at 1	-	-	-	-	1	332 (217)	0.91 (0.87, 0.95)	0.70 (0.61, 0.79)
Menzies for amelanotic at 0	-	-	-	-	1	332 (217)	1.00 (0.98, 1.00)	0.52 (0.43, 0.62)
c. Any skin cancer or lesion with high risk of progression to melanoma	**IN-PERSON**				**IMAGE-BASED**			
No algorithm at NR	1	231 (77)	0.90 (0.81, 0.95)	0.94 (0.89, 0.97)	2	83 (32)	0.78 (0.61, 0.89)	0.75 (0.61, 0.85)
pattern analysis - threshold NR	1	3372 (98)	0.90 (0.82, 0.95)	1.00 (0.99, 1.00)	1	119 (37)	1.00 (0.91, 1.00)	0.96 (0.90, 0.99)
ABCD at > 5.45	1	200 (46)	0.98 (0.88, 1.00)	0.98 (0.94, 1.00)	-	-	-	-
3PCL at >= 2	1	77 (39)	0.85 (0.69, 0.94)	0.26 (0.13, 0.43)	1	150 (44)	0.91 (0.78, 0.97)	0.72 (0.62, 0.80)

Footnotes

All analyses by algorithm were undertaken using the bivariate normal model (BVN).

CI - confidence interval; FU - follow-up; NR - not reported; ABCD(E) - asymmetry, border, colour, differential structures (enlargement); 7FFM - seven features for melanoma; 7PCL - seven point checklist; rev - revised; 3PCL - three point checklist; CASH - colour, architecture, symmetry and homogeneity

6 Analysis by observer experience for detection of invasive melanoma or intraepidermal melanocytic variants

Test	Studies	Lesions (Cases)	DOR (95% CI)	Specificity at 80% sensitivity (95% CI)	Sensitivity at 80% specificity (95% CI)	Relative DOR (95% CI)	P-value (LR)
Experience in-person							

NR	10	8390 (1015)	97.7 (35.6, 268)	97% (90, 99)	94% (87, 98)	1.9 (0.49, 7.1)	0.64
High	14	14213 (612)	52.4 (21.6, 127)	94% (84, 98)	91% (83, 96)	1.00 (comparator)	
Trained	2	566 (37)	19.2 (1.6, 226)	82% (19, 99)	82% (36, 97)	0.37 (0.03, 5.1)	
Experience image-based							
NR	11	2777 (465)	35.4 (15.9, 78.7)	90% (80, 96)	89% (80, 95)	2.0 (0.8, 4.9)	<0.001
High	34	8933 (1956)	17.2 (11.8, 26.5)	82% (74, 87)	81% (75, 86)	1.00 (comparator)	
Moderate	5	678 (193)	11.3 (5.9, 21.3)	73% (58, 85)	74% (61, 84)	0.64 (0.37, 1.1)	
Low	6	448 (123)	5.3 (2.6, 10.8)	55% (35, 73)	58% (41, 74)	0.30 (0.15, 0.58)	
Mixed	5	473 (117)	4.4 (1.4, 13.5)	50% (23, 77)	54% (29, 78)	0.25 (0.07, 0.81)	
Trained	11	1087 (240)	9.0 (4.5, 17.9)	68% (51, 81)	70% (55, 82)	0.15 (0.25, 1.02)	

Footnotes

DOR - diagnostic odds ratio; CI - confidence interval; NR - not reported; LR - likelihood ratio test

7 Analysis by observer qualifications for detection of invasive melanoma or intraepidermal melanocytic variants

Test	Studies	Lesions (Cases)	DOR (95% CI)	Specificity at 80% sensitivity (95% CI)	Sensitivity at 80% specificity (95% CI)	Relative DOR (95% CI)	P value (LR)
Qualifications in-person							
Consultant expert	11	2767 (439)	52.4 (21.6, 127)	94% (84, 98)	91% (83, 96)	1.00 (comparator)	0.33
Consultant	10	8390 (1015)	97.7 (35.6, 268)	97% (90, 99)	95% (87, 98)	1.86 (0.949, 7.11)	
GP	2	566 (37)	19.2 (1.6, 226)	82% (19, 99)	82% (36, 97)	0.37 (0.03, 5.08)	
Resident/registrar(1)	2	11137 (127)	51.6 (2.9, 927)	93% (42, 100)	93% (42, 100)	not estimable within model	-
Mixed (secondary care)(1)	1	309 (46)	29.6 (13.5, 64.8)	88% (77, 94)	88% (77, 94)	not estimable within model	-
Qualifications image based							
Consultant expert	33	8664 (1854)	19.4 (13.1,28.8)	83% (76, 88)	83% (77, 88)	1.0 (comparator)	<0.001 -
Consultant	25	4589 (955)	11.9 (7.6, 18.6)	74% (65, 82)	75% (66, 82)	0.61 (0.40, 0.92)	
Resident	5	927 (138)	6.0 (2.6, 14.0)	59% (37, 78)	61% (41, 78)	0.31 (0.14, 0.71)	
Mixed (other)	4	867 (229)	15.1 (4.0, 57.0)	79% (48, 94)	79% (,)	0.78 (0.20, 3.1)	
GP / Primary care	3	288 (55)	1.9 (0.7, 5.0)	30% (12, 57)	34% (51, 93)	0.10 (0.04, 0.25)	
Mixed (secondary care)(2)	4	399 (111)	10.3 (3.0, 35.3)	72% (43, 90)	72% (43, 90)	not estimable within model	-
Physician assistant(2)	1	65 (25)	3.6 (1.1, 11.5)	47% (22, 74)	47% (22, 74)	not estimable within model	-

Footnotes

DOR - diagnostic odds ratio; CI - confidence interval; NR - not reported; LR - likelihood ratio test

(1) In-person model could not be fitted including the small number of studies in these groups. Estimates for these groups are obtained from computed the DOR for the individual study, or random effects meta-analyses of DORs where there is more than one study. Estimates at the 80% sensitivity and specificity values are computed assuming symmetric SROC curves.

(2) Image-based model could not be fitted including the small number of studies in these groups. Estimates for these groups are obtained from computed the DOR for the individual study, or random effects meta-analyses of DORs where there is more than one study. Estimates at the 80% sensitivity and specificity values are computed assuming symmetric SROC curves.

8 Sensitivity analyses for in-person visual inspection and dermoscopy added to visual inspection for the detection of invasive melanoma or atypical intraepidermal melanocytic variants

Test	Studies	Lesions (cases)	DOR (95% CI)	Specificity at 80% sensitivity (95% CI)	Sensitivity at 80% specificity (95% CI)	Relative DOR (95% CI)
All in-person evaluations						
Visual inspection	13	6740 (459)	13.1 (7.0, 24.5)	75% (57, 87)	76% (66, 85)	4.7 (3.0, 7.5)

Visual inspection + Dermoscopy	26	23169 (1664)	61.7 (34.9, 109)	95% (90, 98)	92% (87, 95)	
Change with adding dermoscopy to VI (95% CI)				+20% (+7, +33)	+16% (+8, +23)	
In-person evaluations (direct comparison)						
Visual inspection	11	5854 (412)	13.7 (5.9, 31.8)	75% (49, 90)	77% (63, 87)	4.8 (2.8, 8.1)
Visual inspection + Dermoscopy	11	5854 (412)	65.7 (27.0, 160)	95% (87, 99)	92% (84, 96)	
Change with adding dermoscopy to VI (95% CI)				+21% (+2, +39)	+15% (+7, +23)	
In-person evaluations (with histology and follow-up for those not having surgery)						
Visual inspection	2	3607 (60)	18.4 (2.63, 128)	82% (39, 97)	82% (40, 97)	14.4 (4.4, 47.6)
Visual inspection + Dermoscopy	6	17574 (800)	265 (49, 1428)	99% (91, 100)	98% (87, 100)	
Change with adding dermoscopy to VI (95% CI)				+16% (-23, 56)	+16% (-20, 53)	
In-person evaluations with low risk of bias for the index test						
Visual inspection	4	3957 (176)	16.9 (6.1, 46.8)	80% (52, 94)	80% (63, 91)	3.1 (1.3, 7.4)
Visual inspection + Dermoscopy	20	19182 (831)	53.0 (25.8, 109)	94% (87, 98)	91% (84,95)	
Change with adding dermoscopy to VI (95% CI)				+14% (-6, +34)	+11% (-1, +23)	
In-person evaluations with low risk of bias for the reference test						
Visual inspection	10	2802 (367)	13.8 (7.3, 26.3)	76% (59, 87)	77% (67, 85)	4.2 (2.5, 7.1)
Visual inspection + Dermoscopy	20	7636 (1418)	57.8 (32.2, 104)	95% (89, 97)	92% (87, 95)	
Change with adding dermoscopy to VI (95% CI)				+19% (+5, +32)	+15% (+6, +23)	
In-person evaluations with low risk of bias for flow and timing						
Visual inspection	2	601 (66)	11.0 (2.7, 44.4)	61% (26, 87)	73% (55, 85)	5.1 (1.2, 20.9)
Visual inspection + Dermoscopy	4	984 (113)	55.7 (24.4, 127)	95% (85, 98)	88% (79, 94)	
Change with adding dermoscopy to VI (95% CI)				+34% (-45, +100)	+16% (-28, +60)	
In-person evaluations excluding case-control studies						
Visual inspection	13	6740 (459)	13.1 (7.0, 24.5)	75% (57, 87)	76% (66, 85)	4.7 (3.0, 7.5)
Visual inspection + Dermoscopy	26	23169 (1664)	61.7 (34.9, 109)	95% (90, 98)	92% (87, 95)	
Change with adding dermoscopy to VI (95% CI)				+20% (+7, +33)	+16% (+8, +23)	

Footnotes

DOR - diagnostic odds ratio; CI - confidence interval; NR - not reported

9 Sensitivity analyses for image-based visual inspection or dermoscopy for the detection of invasive melanoma or atypical intraepidermal melanocytic variants

Test	Studies	Lesions (cases)	DOR (95% CI)	Specificity at 80% sensitivity (95% CI)	Sensitivity at 80% specificity (95% CI)	Relative DOR (95% CI)
All Image based evaluations						
Clinical (macro) images	11	1740 (305)	3.2 (1.9, 5.4)	42% (28, 58)	47% (34, 59)	5.6 (3.7, 8.5)
Dermoscopic images	60	13,475 (2851)	17.8 (12.3, 25.7)	82% (75, 87)	81% (76, 86)	
Change replacing VI with dermoscopy (95% CI)				+40% (+27, 57)	+35% (+24, +46)	
Image based evaluations (direct studies)						
Clinical (macro) images	11	1740 (305)	3.6 (1.7, 7.6)	48% (25, 73)	47% (30, 64)	5.3 (3.5, 8.0)
Dermoscopic images	11	1735 (306)	19.2 (8.7,42.0)	83% (70, 91)	83% (68, 92)	
Change replacing VI with dermoscopy (95% CI)				+34% (+15, +53)	+36% (+20, +52)	
Image based evaluations (with histology and follow-up for those not having surgery)						

Clinical (macro) images	0	-	-	-	-	-
Dermoscopic images	7	2612 (523)	7.4 (4.5, 12.0)	67% (58, 75)	57% (39, 74)	
Change replacing VI with dermoscopy (95% CI)				-	-	
Image based evaluations with low risk of bias for the index test						
Clinical (macro) images	3	1113 (117)	1.9 (0.91, 4.0)	+32% (17, 52)	32% (17, 52)	10.4 (5.7, 19.0)
Dermoscopic images	40	11194 (2318)	19.8 (12.4, 31.7)	83% (76, 89)	83% (76, 89)	
Change replacing VI with dermoscopy (95% CI)				+51% (+35, +68)	+51% (+34, +67)	
Image based evaluations with low risk of bias for the reference test						
Clinical (macro) images	9	1650 (276)	3.2 (1.9, 5.4)	42% (28, 58)	47% (34, 59)	5.6 (3.7, 8.5)
Dermoscopic images	51	10894 (2359)	17.8 (12.3, 25.8)	82% (75, 87)	81% (76, 86)	
Change replacing VI with dermoscopy (95% CI)				+40% (+26, +53)	+35% (+24, +46)	
Image based evaluations with low risk of bias for flow and timing						
Clinical (macro) images	1	53 (10)	15.9 (1.6, 161)	79% (21, 98)	80% (34, 97)	0.54 (0.05, 5.5)
Dermoscopic images	11	1391 (410)	8.6 (4.4, 16.7)	65% (42, 83)	69% (56, 80)	
Change replacing VI with dermoscopy (95% CI)				-14% (-67, +39)	-10% (-48, +28)	
Image based evaluations excluding case-control studies						
Clinical (macro) images	7	964 (183)	7.2 (3.5, 14.8)	62% (40, 80)	66% (50, 78)	3.4 (1.8, 6.4)
Dermoscopic images	37	10,270 (1923)	24.3 (15.2, 39.0)	86% (79, 91)	85% (79, 90)	
Change replacing VI with dermoscopy (95% CI)				+24% (+4, +44)	+20% (+7, +32)	

Footnotes

10 Comparison of visual inspection and dermoscopy for the detection of invasive melanoma

Test	Studies	Lesions (cases)	DOR (95% CI)	Specificity at 80% sensitivity	Sensitivity at 80% specificity	Relative DOR (95% CI)	P value (LR)
In-person evaluations							
Visual inspection	2	147 (51)	20.8 (6.0, 72.5)	84% (66, 93)	84% (57, 95)	6.2 (1.5, 26.6)	0.015
Visual inspection +Dermoscopy	6	789 (115)	129 (19.2, 870)	97% (94, 98)	97% (46, 100)		
Difference (95% CI)				+13% (-1, +27)	+13% (-0, +27)		
In-person evaluations (direct studies)							
Visual inspection	2	147 (51)	20.1 (4.0, 101)	75% (23, 97)	78% (64, 88)	11.3 (1.4, 689.8)	0.015
Visual inspection +Dermoscopy	2	147 (51)	396226 (21.7, 2358)	99% (54, 100)	94% (72, 99)		
Difference (95% CI)				+24% (-21, +69)	+15% (+2, +29)		
Image based evaluations							
Clinical (macro) images	4	454 (145)	11.0 (4.1, 29.3)	74% (52, 88)	72% (49, 88)	2.5 (1.2, 5.1)	0.032
Dermoscopic images	13	5618 (1092)	27.5 (12.2, 61.7)	87% (75, 94)	88% (75, 94)		
Difference (95% CI)				+13% (-1, +28)	+15% (-1, +30)		
Image based evaluations (direct studies)							
Clinical (macro) images	4	454 (145)	11.9 (3.4, 40.9)	45% (5, 92)	72% (59, 82)	3.4 (1.0, 11.1)	0.049
Dermoscopic images	4	454	40.4	89%	83%		

		(145)	(8.2, 198)	(47, 99)	(72, 90)		
Difference (95% CI)				+44% (-20, +100)	+11% (+1, +22)		

Footnotes

DOR - diagnostic odds ratio; CI - confidence interval; LR - likelihood ratio test

11 Comparison of visual inspection and dermoscopy for the detection of any skin lesion requiring excision

Test	Studies	Lesions (cases)	DOR (95% CI)	Specificity at 80% sensitivity	Sensitivity at 80% specificity
In-person evaluations					
Visual inspection	2	3457 (151)	38.4 (2.5, 582)	91% (39, 99)	91% (39, 99)
Visual inspection +Dermoscopy	4	3880 (260)	232 (16.0, 3354)	98% (80, 100)	98% (80,100)
In-person evaluations (direct studies)					
Visual inspection	2	3457 (151)	15.0 (0.18, 1225)	79% (4,100)	79% (4,100)
Visual inspection +Dermoscopy	2	3449 (137)	88.1 (1.1, 7338)	96% (21,100)	96% (21,100)
Image-based evaluations					
Clinical (macro) images	3	547 (138)	21.7 (4.8, 98.9)	84% (54, 96)	84% (54, 96)
Dermoscopic images	6	815 (217)	37.5 (8.8, 161)	90% (69, 98)	90% (69, 98)
Image based evaluations (direct studies)					
Clinical (macro) images	3	547 (138)	12.1 (5.4, 26.7)	75% (58, 87)	75% (58, 87)
Dermoscopic images	3	546 (136)	18.4 (8.1, 41.7)	82% (67, 91)	82% (67, 91)

Footnotes

DOR - diagnostic odds ratio; CI - confidence interval

Estimates are based on fitting models with symmetric ROC curves, and no formal comparisons between tests are made due to paucity of data. It is noted that the estimates for the visual inspection studies change between the all data and paired data analyses for both in-person and image-based analyses. This is driven by differences in the heterogeneity in accuracy between the models which affects all parameters in the analyses.

12 Accuracy of dermoscopy before vs after dermoscopy training (all image-based)

Test	Studies	Lesions (cases)	DOR (95% CI)	Specificity at 80% sensitivity (95% CI)	Sensitivity at 80% specificity (95% CI)	Relative DOR (95% CI)	P-value (LR)
Detection of invasive melanoma or atypical intraepidermal melanocytic variants							
Before training	4	245 (65)	6.3 (1.68, 23.5)	62% (27, 88)	60% (30, 84)	1.4 (0.38, 5.3)	<0.001
After training	4	245 (65)	8.9 (2.4, 33.3)	69% (40, 88)	69% (33, 91)		
Change with training (95% CI)				+8% (-24, +40)	+8% (-19, +36)		
Detection of invasive melanoma							
Before training	2	150 (75)	5.2 (0.95, 28.7)	50% (9, 91)	60% (25, 87)	3.2 (0.94, 10.6)	0.051
After training	2	150 (75)	16.4 (2.6, 103)	80% (32, 97)	80% (47, 95)		
Change with training (95% CI)				+29% (-24, +82)	+20% (-5, +45)		

Footnotes

DOR - diagnostic odds ratio; CI - confidence interval; LR - likelihood ratio test

References to studies

Included studies

Ahnlide 2016

* Ahnlide I, Bjellerup M, Nilsson F, Nielsen K. Validity of ABCD Rule of Dermoscopy in Clinical Practice. Acta Dermato-Venereologica 2016;96(3):367-72. [Other: ER4:25012370; PubMed: 26351008]

Alarcon 2014

* Alarcon I, Carrera C, Palou J, Alos L, Malvehy J, Puig S. Impact of in vivo reflectance confocal microscopy on the number needed to treat melanoma in doubtful lesions. British Journal of Dermatology 2014;170(4):802-8. [Other: ER4:17941078; PubMed: 24124911]

Annessi 2007

* Annessi G, Bono R, Sampogna F, Faraggiana T, Abeni D. Sensitivity, specificity, and diagnostic accuracy of three dermoscopic algorithmic methods in the diagnosis of doubtful melanocytic lesions: the importance of light brown structureless areas in differentiating atypical melanocytic nevi from thin melanomas. Journal of the American Academy of Dermatology 2007;56(5):759-67. [Other: ER4:15465846; PubMed: 17316894]

Arevalo 2008

* Arevalo A, Altamura D, Avramidis M, Blum A, Menzies S. The significance of eccentric and central hyperpigmentation, multifocal hyper/hypopigmentation, and the multicomponent pattern in melanocytic lesions lacking specific dermoscopic features of melanoma. Archives of Dermatology 2008;144(11):1440-4. [Other: ER4:19728335; PubMed: 19015418]

Argenziano 1998

* Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Sammarco E, Delfino M. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Archives of Dermatology 1998;134(12):1563-70. [Other: ER4:15465850; PubMed: 9875194]

Argenziano 2006

* Argenziano G, Puig S, Zalaudek I, Sera F, Corona R, Alsina M, et al. Dermoscopy improves accuracy of primary care physicians to triage lesions suggestive of skin cancer. Journal of Clinical Oncology 2006;24(12):1877-82. [Other: ER4:17940973; PubMed: 16622262]

Argenziano 2011

* Argenziano G, Catricala C, Ardigo M, Buccini P, De Simone P, Eibenschutz L, et al. Seven-point checklist of dermoscopy revisited. British Journal of Dermatology 2011;164(4):785-90. [Other: ER4:15465848; PubMed: 21175563]

Ascierto 2010

* Ascierto PA, Palla M, Ayala F, De Michele I, Caraco C, Daponte A, et al. The role of spectrophotometry in the diagnosis of melanoma. BMC Dermatology 2010;10:5. [Other: ER4:19728329; PubMed: 20707921]

Bauer 2000

* Bauer P, Cristofolini P, Boi S, Burroni M, Dell'Eva G, Micciolo R, et al. Digital epiluminescence microscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. A statistical comparison between visual and computer inspection. Melanoma Research 2000;10(4):345-9. [Other: ER4:15465861; PubMed: 10985668]

Benelli 1999

* Benelli C, Roscetti E, Pozzo VD, Gasparini G, Cavicchini S. The dermoscopic versus the clinical diagnosis of melanoma. European Journal of Dermatology 1999;9(6):470-6. [Other: ER4:18375029; PubMed: 10491506]

Benelli 2000

Benelli C, Roscetti E, Dal Pozzo V. The dermoscopic (7FFM) versus the clinical (ABCDE) diagnosis of small diameter melanoma. European Journal of Dermatology 2000;10(4):282-7. [PubMed: 10846255]

Benelli 2001

* Benelli C, Roscetti E, Dal Pozzo V. Reproducibility of the clinical criteria (ABCDE rule) and dermatoscopic features (7FFM) for the diagnosis of malignant melanoma. European Journal of Dermatology 2001;11(3):234-9. [Other: ER4:18375028; PubMed: 11358731]

Binder 1994

* Binder M, Steiner A, Schwarz M, Knollmayer S, Wolff K, Pehamberger H. Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study. British Journal of Dermatology 1994;130(4):460-5. [Other: ER4:18375032; PubMed: 8186110]

Binder 1995

* Binder M, Schwarz M, Winkler A, Steiner A, Kaider A, Wolff K, et al. Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. Archives of Dermatology 1995;131(3):286-91. [Other: ER4:18375031; PubMed: 7887657]

Binder 1999

* Binder M, Kittler H, Steiner A, Dawid M, Pehamberger H, Wolff K. Reevaluation of the ABCD rule for epiluminescence microscopy. Journal of the American Academy of Dermatology 1999;40(2 Pt 1):171-6. [Other: ER4:15465864; PubMed: 10025741]

Blum 2003

* Blum A, Rassner G, Garbe C. Modified ABC-point list of dermoscopy: A simplified and highly accurate dermoscopic algorithm for the diagnosis of cutaneous melanocytic lesions. Journal of the American Academy of Dermatology 2003;48(5):672-8. [Other: ER4:15465867; PubMed: 12734495]

Blum 2003b

* Blum A, Soyer HP, Garbe C, Kerl H, Rassner G, Hofmann-Wellenhof R. The dermoscopic classification of atypical melanocytic naevi (Clark naevi) is useful to discriminate benign from malignant melanocytic lesions. British Journal of Dermatology 2003;149(6):1159-64. [Other: ER4:15465868; PubMed: 14674892]

Blum 2004

* Blum A, Hofmann-Wellenhof R, Luedtke H, Ellwanger U, Steins A, Roehm S, et al. Value of the clinical history for different users of dermoscopy compared with results of digital image analysis. Journal of the European Academy of Dermatology & Venereology 2004;18(6):665-9. [Other: ER4:15465865; PubMed: 15482291]

Blum 2004b

* Blum A, Luedtke H, Ellwanger U, Schwabe R, Rassner G, Garbe C. Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. British Journal of Dermatology 2004;151(5):1029-38. [Other: ER4:15465866; PubMed: 15541081]

Bono 2002

* Bono A, Bartoli C, Cascinelli N, Lualdi M, Maurichi A, Moglia D, et al. Melanoma detection. A prospective study comparing diagnosis with the naked eye, dermatoscopy and telespectrophotometry. Dermatology 2002;205(4):362-6. [Other: ER4:15465870; PubMed: 12444332]

Bono 2002b

* Bono A, Bartoli C, Baldi M, Tomatis S, Bifulco C, Santinami M. Clinical and dermatoscopic diagnosis of small pigmented skin lesions. European Journal of Dermatology 2002;12(6):573-6. [Other: ER4:18375034; PubMed: 12459531]

Bono 2006

* Bono A, Tolomio E, Trincone S, Bartoli C, Tomatis S, Carbone A, et al. Micro-melanoma detection: a clinical study on 206 consecutive cases of pigmented skin lesions with a diameter < or = 3 mm. British Journal of Dermatology 2006;155(3):570-3. [Other: ER4:15465872; PubMed: 16911283]

Bourne 2012

* Bourne P, Rosendahl C, Keir J, Cameron A. BLINCK-A diagnostic algorithm for skin cancer diagnosis combining clinical features with dermatoscopy findings. Dermatology Practical & Conceptual 2012;2(2):202a12. [Other: ER4:17941081; PubMed: 23785600]

Broganelli 2005

* Broganelli P, Chiaretta A, Sacerdote C, Pippione M. The epiluminescence microscopy in the ambulatory clinical practice: Diagnostic accuracy and usefulness of videodermatoscopic monitoring [L'epiluminescenza nella pratica clinica ambulatoriale: Accuratezza diagnostica ed utilita del monitoraggio videodermatoscopico]. Giornale Italiano di Dermatologia e Venereologia 2005;140(1):15-25. [Other: ER4:18375073]

Carli 1994

* Carli P, De Giorgi V, Donati E, Pestelli E, Giannotti B. Epiluminescence microscopy reduces the risk of removing clinically atypical, but histologically common, melanocytic lesions [La Microscopia a Epiluminescenza (Elm) Riduce Il Rischio Di Asportare Lesioni Melanocitarie Clinicamente Sospette Ma Istologicamente Comuni]. Giornale Italiano di Dermatologia e Venereologia 1994;129(12):599-605. [Other: ER4:18375075]

Carli 2002

* Carli P, De Giorgi V, Argenziano G, Palli D, Giannotti B. Pre-operative diagnosis of pigmented skin lesions: in vivo dermoscopy performs better than dermoscopy on photographic images. Journal of the European Academy of Dermatology & Venereology 2002;16(4):339-46. [Other: ER4:15465882; PubMed: 12224689]

Carli 2002b

* Carli P, de Giorgi V, Salvini C, Mannone F, Chiarugi A. The gold standard for photographing pigmented skin lesions for diagnostic purposes: contact versus distant imaging. Skin Research & Technology 2002;8(4):255-9. [Other: ER4:15465888; PubMed: 12423545]

Carli 2003

* Carli P, Quercioli E, Sestini S, Stante M, Ricci L, Brunasso G, et al. Pattern analysis, not simplified algorithms, is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology. British Journal of Dermatology 2003;148(5):981-4. [Other: ER4:15465890; PubMed: 12786829]

Carli 2003b

* Carli P, De Giorgi V, Chiarugi A, Nardini P, Mannone F, Stante M, et al. Effect of lesion size on the diagnostic performance of dermoscopy in melanoma detection. Dermatology 2003;206(4):292-6. [Other: ER4:15465883; PubMed: 12771468]

Carrera 2016

* Carrera C, Marchetti MA, Dusza SW, Argenziano G, Braun RP, Halpern AC, et al. Validity and reliability of dermoscopic criteria used to differentiate nevi from melanoma a web-based international dermoscopy society study. JAMA Dermatology 2016;152(7):798-806. [Other: ER4:25233595; PubMed: 27074267]

Coras 2003

Coras B, Glaessl A, Kinateder J, Klovekorn W, Braun R, Lepski U, et al. Teledermatoscopy in daily routine--results of the first 100 cases. Current Problems in Dermatology 2003;32:207-12. [PubMed: 12472014]

Cristofolini 1994

* Cristofolini M, Zumiani G, Bauer P, Cristofolini P, Boi S, Micciolo R. Dermatoscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. Melanoma Research 1994;4(6):391-4. [Other: ER4:15465898; PubMed: 7703719]

Dal Pozzo 1999

* Dal Pozzo V, Benelli C, Roscetti E. The seven features for melanoma: a new dermoscopic algorithm for the diagnosis of malignant melanoma. European Journal of Dermatology 1999;9(4):303-8. [Other: ER4:18375041; PubMed: 10356410]

di Meo 2016

* di Meo N, Stinco G, Bonin S, Gatti A, Trevisini S, Damiani G, et al. CASH algorithm versus 3-point checklist and its modified version in evaluation of melanocytic pigmented skin lesions: The 4-point checklist. Journal of Dermatology 2016;43(6):682-5. [Other: ER4:25012343; PubMed: 26589251]

Dolianitis 2005

* Dolianitis C, Kelly J, Wolfe R, Simpson P. Comparative performance of 4 dermoscopic algorithms by nonexperts for the diagnosis of melanocytic lesions. Archives of Dermatology 2005;141(8):1008-14. [Other: ER4:15465906; PubMed: 16103330]

Dreiseitl 2009

* Dreiseitl S, Binder M, Hable K, Kittler H. Computer versus human diagnosis of melanoma: evaluation of the feasibility of an automated diagnostic system in a prospective clinical trial. Melanoma Research 2009;19(3):180-4. [Other: ER4:15465907; PubMed: 19369900]

Duff 2001

* Duff CG, Melsom D, Rigby HS, Kenealy JM, Townsend PL. A 6 year prospective analysis of the diagnosis of malignant melanoma in a pigmented-lesion clinic: even the experts miss malignant melanomas, but not often. British Journal of Plastic Surgery 2001;54(4):317-321. [DOI: http://dx.doi.org/10.1054/bjps.2000.3561; Other: ER4:20569450; PubMed: 11355986]

Dummer 1993

* Dummer W, Doehnel KA, Remy W. [Videomicroscopy in differential diagnosis of skin tumors and secondary prevention of malignant melanoma]. Hautarzt 1993;44(12):772-6. [Other: ER4:18375044; PubMed: 8113040]

Durdu 2011

* Durdu M, Baba M, Seckin D. Dermatoscopy versus Tzanck smear test: a comparison of the value of two tests in the diagnosis of pigmented skin lesions. Journal of the American Academy of Dermatology 2011;65(5):972-82. [Other: ER4:15465910; PubMed: 21565420]

Feci 2015

* Feci L, Cevenini G, Nami N, Fagiolini A, Perotti R, Miracco C, et al. Influence of Ambient Stressors and Time Constraints on Diagnostic Accuracy of Borderline Pigmented Skin Lesions. Dermatology 2015;231(3):269-73. [Other: ER4:25012339; PubMed: 26375805]

Feldmann 1998

* Feldmann R, Fellenz C, Gschnait F. The ABCD rule in dermatoscopy: analysis of 500 melanocytic lesions. Hautarzt 1998;49(6):473-6. [Other: ER4:15465916; PubMed: 9675574]

Ferrari 2015

* Ferrari B, Pupelli G, Farnetani F, De Carvalho NT, Longo C, Reggiani C, et al. Dermoscopic difficult lesions: an objective evaluation of reflectance confocal microscopy impact for accurate diagnosis. Journal of the European Academy of Dermatology and Venereology 2015;29(6):1135-1140. [DOI: 10.1111/jdv.12769; Other: ER4:20569458; PubMed: 25303304]

Ferris 2015

* Ferris LK, Harkes JA, Gilbert B, Winger DG, Golubets K, Akilov O, et al. Computer-aided classification of melanocytic lesions using dermoscopic images. Journal of the American Academy of Dermatology 2015;73(5):769-76. [Other: ER4:25012337; PubMed: 26386631]

Friedman 2008

* Friedman RJ, Gutkowicz-Krusin D, Farber MJ, Warycha M, Schneider-Kels L, Papastathis N, et al. The diagnostic performance of expert dermoscopists vs a computer-vision system on small-diameter melanomas. Archives of Dermatology 2008;144(4):476-82. [Other: ER4:15465921; PubMed: 18427041]

Gereli 2010

* Gereli MC, Onsun N, Atilganoglu U, Demirkesen C. Comparison of two dermoscopic techniques in the diagnosis of clinically atypical pigmented skin lesions and melanoma: seven-point and three-point checklists. International Journal of Dermatology 2010;49(1):33-8. [Other: ER4:15465929; PubMed: 20465608]

Gilmore 2010

* Gilmore S, Hofmann-Wellenhof R, Soyer HP. A support vector machine for decision support in melanoma recognition. Experimental Dermatology 2010;19(9):830-5. [Other: ER4:15465935; PubMed: 20629732]

Glud 2009

* Glud M, Gniadecki R, Drzewiecki KT. Spectrophotometric intracutaneous analysis versus dermoscopy for the diagnosis of pigmented skin lesions: prospective, double-blind study in a secondary reference centre. Melanoma Research 2009;19(3):176-9. [Other: ER4:18375045; PubMed: 19319002]

Gokdemir 2011

* Gokdemir A, Guler OM, Bek Y, Aydin F, Senturk N, Canturk T, et al. Dermoscopic and histopathological correlation in melanocytic and non-melanocytic lesions [Melanositik ve non-melanositik lezyonlarda dermoskopik ve histopatolojik tani korelasyonu]. Turkiye Klinikleri Dermatoloji 2011;21(1):7-16. [EMBASE: 361807346; Other: ER4:18375084]

Grimaldi 2009

* Grimaldi L, Silvestri A, Brandi C, Nisi G, Brafa A, Calabro M, et al. Digital epiluminescence dermoscopy for pigmented cutaneous lesions, primary care physicians, and telediagnosis: a useful tool? Journal of Plastic, Reconstructive & Aesthetic Surgery: JPRAS 2009;62(8):1054-8. [Other: ER4:15465940; PubMed: 18547883]

Guitera 2009a (Modena)

Guitera P, Pellacani G, Longo C, Seidenari S, Avramidis M, Menzies SW. In vivo reflectance confocal microscopy enhances secondary evaluation of melanocytic lesions. Journal of Investigative Dermatology 2009;129(1):131-8. [PubMed: 18633444]

Guitera 2009b (Sydney)

* Guitera P, Pellacani G, Longo C, Seidenari S, Avramidis M, Menzies SW. In vivo reflectance confocal microscopy enhances secondary evaluation of melanocytic lesions. Journal of Investigative Dermatology 2009;129(1):131-8. [Other: ER4:15465945; PubMed: 18633444]

Haenssle 2010a (FV)

Haenssle HA, Korpas B, Hansen-Hagge C, Buhl T, Kaune KM, Rosenberger A, et al. Seven-point checklist for dermatoscopy: performance during 10 years of prospective surveillance of patients at increased melanoma risk. Journal of the American Academy of Dermatology 2010;62(5):785-93. [PubMed: 20226567]

Haenssle 2010b (FU)

Haenssle HA, Korpas B, Hansen-Hagge C, Buhl T, Kaune KM, Rosenberger A, et al. Seven-point checklist for dermatoscopy: performance during 10 years of prospective surveillance of patients at increased melanoma risk. Journal of the American Academy of Dermatology 2010;62(5):785-93. [PubMed: 20226567]

Hauschild 2014

* Hauschild A, Chen SC, Weichenthal M, Blum A, King HC, Goldsmith J, et al. To excise or not: impact of MelaFind on German dermatologists' decisions to biopsy atypical lesions. Journal der Deutschen Dermatologischen Gesellschaft 2014;12(7):606-14. [Other: ER4:17941085; PubMed: 24944011]

Kittler 1998

* Kittler H, Seltenheim M, Pehamberger H, Wolff K, Binder M. Diagnostic informativeness of compressed digital epiluminescence microscopy images of pigmented skin lesions compared with photographs. Melanoma Research 1998;8(3):255-60. [Other: ER4:17941060; PubMed: 9664147]

Kittler 1999

* Kittler H, Seltenheim M, Dawid M, Pehamberger H, Wolff K, Binder M. Morphologic changes of pigmented skin lesions: a useful extension of the ABCD rule for dermatoscopy. Journal of the American Academy of Dermatology 1999;40(4):558-62. [Other: ER4:15465976; PubMed: 10188673]

Kittler 2001

* Kittler H, Binder M. Risks and benefits of sequential imaging of melanocytic skin lesions in patients with multiple atypical nevi. Archives of Dermatology 2001;137(12):1590-5. [Other: ER4:20569472; PubMed: 11735709]

Krahn 1998

* Krahn G, Gottlober P, Sander C, Peter RU. Dermatoscopy and high frequency sonography: two useful non-invasive methods to increase preoperative diagnostic accuracy in pigmented skin lesions. Pigment Cell Research 1998;11(3):151-4. [Other: ER4:15465981; PubMed: 9730322]

Kreusch 1992

Kreusch J, Rassner G, Trahn C, Pietsch-Breitfeld B, Henke D, Selbmann HK. Epiluminescent microscopy: a score of morphological features to identify malignant melanoma. Pigment Cell Research 1992;Suppl 2:295-8. [PubMed: 1409432]

Langley 2007

* Langley RG, Walsh N, Sutherland AE, Propperova I, Delaney L, Morris SF, et al. The diagnostic accuracy of in vivo confocal scanning laser microscopy compared to dermoscopy of benign and malignant melanocytic lesions: a prospective study. Dermatology 2007;215(4):365-72. [Other: ER4:15465985; PubMed: 17912001]

Lorentzen 1999

* Lorentzen H, Weismann K, Petersen CS, Larsen FG, Secher L, Skodt V. Clinical and dermatoscopic diagnosis of malignant melanoma. Assessed by expert and non-expert groups. Acta Dermato-Venereologica 1999;79(4):301-4. [Other: ER4:17941062; PubMed: 10429989]

Lorentzen 2000

* Lorentzen H, Weismann K, Kenet RO, Secher L, Larsen FG. Comparison of dermatoscopic ABCD rule and risk stratification in the diagnosis of malignant melanoma. Acta Dermato-Venereologica 2000;80(2):122-6. [Other: ER4:17941061; PubMed: 10877133]

Lorentzen 2008

* Lorentzen HF, Eefsen RL, Weismann K. Comparison of classical dermatoscopy and acrylic globe magnifier dermatoscopy. Acta Dermato-Venereologica 2008;88(2):139-42. [Other: ER4:15465993; PubMed: 18311441]

Malvehy 2014

Malvehy J, Hauschild A, Curiel-Lewandrowski C, Mohr P, Hofmann-Wellenhof R, Motley R, et al. Clinical performance of the Nevisense system in cutaneous melanoma detection: An international, multicentre, prospective and blinded clinical trial on efficacy and safety. British Journal of Dermatology 2014;171(5):1099-107. [PubMed: 24841846]

Menzies 1996

* Menzies SW, Ingvar C, Crotty KA, McCarthy WH. Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features. Archives of Dermatology 1996;132(10):1178-82. [Other: ER4:21450627; PubMed: 8859028]

Menzies 2005

* Menzies SW, Bischof L, Talbot H, Gutenev A, Avramidis M, Wong L, et al. The performance of SolarScan: an automated dermoscopy image analysis instrument for the diagnosis of primary melanoma. Archives of Dermatology 2005;141(11):1388-96. [Other: ER4:20569478; PubMed: 16301386]

Menzies 2008

Menzies SW, Kreusch J, Byth K, Pizzichetta MA, Marghoob A, Braun R, et al. Dermoscopic evaluation of amelanotic and hypomelanotic melanoma. Archives of Dermatology 2008;144(9):1120-7. [PubMed: 18794455]

Menzies 2009

* Menzies SW, Emery J, Staples M, Davies S, McAvoy B, Fletcher J, et al. Impact of dermoscopy and short-term sequential digital dermoscopy imaging for the management of pigmented lesions in primary care: a sequential intervention trial. British Journal of Dermatology 2009;161(6):1270-7. [DOI: 10.1111/j.1365-2133.2009.09374.x; Other: ER4:15466005; PubMed: 19747359]

Menzies 2013

Menzies SW, Moloney FJ, Byth K, Avramidis M, Argenziano G, Zalaudek I, et al. Dermoscopic evaluation of nodular melanoma. JAMA Dermatology 2013;149(6):699-709. [PubMed: 23553375]

Morales Callaghan 2008

* Morales-Callaghan AM, Castrodeza-Sanz J, Martinez-Garcia G, Peral-Martinez I, Miranda-Romero A. Correlation between clinical, dermatoscopic, and histopathologic variables in atypical melanocytic nevi. Actas Dermo-Sifiliograficas 2008;99(5):380-9. [Other: ER4:17941068; PubMed: 18501170]

Nachbar 1994

* Nachbar F, Stolz W, Merkle T, Cognetta AB, Vogt T, Landthaler M, et al. The ABCD rule of dermatoscopy. High prospective value in the diagnosis of doubtful melanocytic skin lesions. Journal of the American Academy of Dermatology 1994;30(4):551-9. [Other: ER4:15466022; PubMed: 8157780]

Nilles 1994

* Nilles M, Boedeker RH, Schill WB. Surface microscopy of naevi and melanomas--clues to melanoma. British Journal of Dermatology 1994;130(3):349-55. [Other: ER4:18375123; PubMed: 8148277]

Pagnanelli 2003

* Pagnanelli G, Soyer HP, Argenziano G, Talamini R, Barbati R, Bianchi L, et al. Diagnosis of pigmented skin lesions by dermoscopy: web-based training improves diagnostic performance of non-experts. British Journal of Dermatology 2003;148(4):698-702. [Other: ER4:15466036; PubMed: 12752126]

Piccolo 2000

Piccolo D, Smolle J, Argenziano G, Wolf IH, Braun R, Cerroni L, et al. Teledermoscopy--results of a multicentre study on 43 pigmented skin lesions. Journal of Telemedicine & Telecare 2000;6(3):132-7. [PubMed: 10912329]

Piccolo 2002

* Piccolo D, Ferrari A, Peris K, Diadone R, Ruggeri B, Chimenti S. Dermoscopic diagnosis by a trained clinician vs. a clinician with minimal dermoscopy training vs. computer-aided diagnosis of 341 pigmented skin lesions: a comparative study. British Journal of Dermatology 2002;147(3):481-6. [Other: ER4:15466057; PubMed: 12207587]

Piccolo 2014

* Piccolo D, Crisman G, Schoinas S, Altamura D, Peris K. Computer-automated ABCD versus dermatologists with different degrees of experience in dermoscopy. European Journal of Dermatology 2014;24(4):477-81. [Other: ER4:17941089; PubMed: 24721784]

Pizzichetta 2002

* Pizzichetta MA, Talamini R, Piccolo D, Trevisan G, Veronesi A, Carbone A, et al. Interobserver agreement of the dermoscopic diagnosis of 129 small melanocytic skin lesions. Tumori 2002;88(3):234-8. [Other: ER4:18375049; PubMed: 12195762]

Pizzichetta 2004

* Pizzichetta MA, Talamini R, Stanganelli I, Puddu P, Bono R, Argenziano G, et al. Amelanotic/hypomelanotic melanoma: clinical and dermoscopic features. British Journal of Dermatology 2004;150(6):1117-24. [Other: ER4:15466066; PubMed: 15214897]

Pupelli 2013

* Pupelli G, Longo C, Veneziano L, Cesinaro AM, Ferrara G, Piana S, et al. Small-diameter melanocytic lesions: morphological analysis by means of in vivo confocal microscopy. British Journal of Dermatology 2013;168(5):1027-33. [Other: ER4:15466070; PubMed: 23301553]

Rao 1997

* Rao BK, Marghoob AA, Stolz W, Kopf AW, Slade J, Wasti Q, et al. Can early malignant melanoma be differentiated from atypical melanocytic nevi by in vivo techniques? Part I. Clinical and dermoscopic characteristics. Skin Research and Technology 1997;3(1):8-14. [EMBASE: 27145858; Other: ER4:17941048]

Rigel 2012

* Rigel DS, Roy M, Yoo J, Cockerell CJ, Robinson JK, White R. Impact of guidance from a computer-aided multispectral digital skin lesion analysis device on decision to biopsy lesions clinically suggestive of melanoma. Archives of Dermatology 2012;148(4):541-3. [Other: ER4:15466080; PubMed: 22351788]

Rosendahl 2011

* Rosendahl C, Tschandl P, Cameron A, Kittler H. Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic pigmented lesions. Journal of the American Academy of Dermatology 2011;64(6):1068-73. [Other: ER4:15466083; PubMed: 21440329]

Rubegni 2012

* Rubegni P, Cevenini G, Nami N, Argenziano G, Saida T, Burroni M, et al. Dermoscopy and digital dermoscopy analysis of palmoplantar 'equivocal' pigmented skin lesions in Caucasians. Dermatology 2012;225(3):248-55. [Other: ER4:15466088; PubMed: 23182753]

Rubegni 2016

* Rubegni P, Tognetti L, Argenziano G, Nami N, Brancaccio G, Cinotti E, et al. A risk scoring system for the differentiation between melanoma with regression and regressing nevi. Journal of Dermatological Science 2016;83(2):138-44. [Other: ER4:25012293; PubMed: 27157925]

Sboner 2004

* Sboner A, Bauer P, Zumiani G, Eccher C, Blanzieri E, Forti S, et al. Clinical validation of an automated system for supporting the early diagnosis of melanoma. Skin Research & Technology 2004;10(3):184-92. [Other: ER4:15466104; PubMed: 15225269]

Seidenari 1998

* Seidenari S, Pellacani G, Pepe P. Digital videomicroscopy improves diagnostic accuracy for melanoma. Journal of the American Academy of Dermatology 1998;39(2 Pt 1):175-81. [Other: ER4:15466116; PubMed: 9704824]

Seidenari 2005

* Seidenari S, Pellacani G, Martella A. Acquired melanocytic lesions and the decision to excise: role of color variegation and distribution as assessed by dermoscopy. Dermatologic Surgery 2005;31(2):184-9. [Other: ER4:15466115; PubMed: 15762212]

Seidenari 2007

* Seidenari S, Grana C, Pellacani G. Colour clusters for computer diagnosis of melanocytic lesions. Dermatology 2007;214(2):137-43. [Other: ER4:15466111; PubMed: 17341863]

Skvara 2005

* Skvara H, Teban L, Fiebiger M, Binder M, Kittler H. Limitations of dermoscopy in the recognition of melanoma. Archives of Dermatology 2005;141(2):155-60. [Other: ER4:20569495; PubMed: 15724011]

Soyer 1995

* Soyer HP, Smolle J, Leitinger G, Rieger E, Kerl H. Diagnostic reliability of dermoscopic criteria for detecting malignant melanoma. Dermatology 1995;190(1):25-30. [Other: ER4:18375054; PubMed: 7894091]

Soyer 2004

* Soyer HP, Argenziano G, Zalaudek I, Corona R, Sera F, Talamini R, et al. Three-point checklist of dermoscopy. A new screening method for early detection of melanoma. Dermatology 2004;208(1):27-31. [Other: ER4:15466124; PubMed: 14730233]

Stanganelli 1998

* Stanganelli I, Serafini M, Cainelli T, Cristofolini M, Baldassari L, Staffa M, et al. Accuracy of epiluminescence microscopy among practical dermatologists: a study from the Emilia-Romagna region of Italy. Tumori 1998;84(6):701-5. [Other: ER4:18375055; PubMed: 10080681]

Stanganelli 1999

* Stanganelli I, Seidenari S, Serafini M, Pellacani G, Bucchi L. Diagnosis of pigmented skin lesions by epiluminescence microscopy: determinants of accuracy improvement in a nationwide training programme for practical dermatologists. Public Health 1999;113(5):237-42. [Other: ER4:15466128; PubMed: 10557118]

Stanganelli 2000

* Stanganelli I, Serafini M, Bucch L. A cancer-registry-assisted evaluation of the accuracy of digital epiluminescence microscopy associated with clinical examination of pigmented skin lesions. Dermatology 2000;200(1):11-6. [Other: ER4:15466129; PubMed: 10681607]

Stanganelli 2005

* Stanganelli I, Brucale A, Calori L, Gori R, Lovato A, Magi S, et al. Computer-aided diagnosis of melanocytic lesions. Anticancer Research 2005;25(6C):4577-82. [Other: ER4:15466126; PubMed: 16334145]

Stanganelli 2015

* Stanganelli I, Longo C, Mazzoni L, Magi S, Medri M, Lanzanova G, et al. Integration of reflectance confocal microscopy in sequential dermoscopy follow-up improves melanoma detection accuracy. British Journal of Dermatology 2015;172(2):365-371. [Other: ER4:20569496; PubMed: 25154446]

Stolz 1994

* Stolz W, Riemann A, Cognetta AB, Pillet L, Abmayer W, Holzel D, et al. ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma. European Journal of Dermatology 1994;4(7):521-527. [EMBASE: 24349113; Other: ER4:18375098]

Tan 2009

* Tan E, Levell NJ. Regular clinical dermatoscope use with training improves melanoma diagnosis by dermatologists. Clinical & Experimental Dermatology 2009;34(8):e876-8. [Other: ER4:17941000; PubMed: 20055853]

Tenenhaus 2010

* Tenenhaus A, Nkengne A, Horn JF, Serruys C, Giron A, Fertil B. Detection of melanoma from dermoscopic images of naevi acquired under uncontrolled conditions. Skin Research & Technology 2010;16(1):85-97. [Other: ER4:17941001; PubMed: 20384887]

Troyanova 2003

* Troyanova P. A beneficial effect of a short-term formal training course in epiluminescence microscopy on the diagnostic performance of dermatologists about cutaneous malignant melanoma. Skin Research & Technology 2003;9(3):269-73. [Other: ER4:17941004; PubMed: 12877690]

Unlu 2014

* Unlu E, Akay BN, Erdem C. Comparison of dermatoscopic diagnostic algorithms based on calculation: The ABCD rule of dermatoscopy, the seven-point checklist, the three-point checklist and the CASH algorithm in dermatoscopic evaluation of melanocytic lesions. Journal of Dermatology 2014;41(7):598-603. [Other: ER4:15466145; PubMed: 24807635]

Viglizzo 2004

* Viglizzo G, Rongioletti F. Clinical, dermoscopic and pathologic correlation of pigmentary lesions observed in a dermoscopy service in the year 2003 [Correlazione clinico-dermoscopico-patologica delle lesioni cutanee pigmentate osservate in un servizio di dermoscopia nell'anno 2003]. Giornale Italiano di Dermatologia e Venereologia 2004;139:339-344. [EMBASE: 39456561; Other: ER4:18375099]

Wells 2012

* Wells R, Gutkowicz-Krusin D, Veledar E, Toledano A, Chen SC. Comparison of diagnostic and management sensitivity to melanoma between dermatologists and MelaFind: a pilot study. Archives of Dermatology 2012;148(9):1083-4. [Other: ER4:15466163; PubMed: 22986873]

Westerhoff 2000

* Westerhoff K, McCarthy WH, Menzies SW. Increase in the sensitivity for melanoma diagnosis by primary care physicians using skin surface microscopy. British Journal of Dermatology 2000;143(5):1016-20. [Other: ER4:15466164; PubMed: 11069512]

Winkelmann 2016

* Winkelmann RR, Farberg AS, Tucker N, White R, Rigel DS. Enhancement of international dermatologists' pigmented skin lesion biopsy decisions following dermoscopy with subsequent integration of multispectral digital skin lesion analysis. Journal of Clinical and Aesthetic Dermatology 2016;9(7):53-55. [Other: ER4:25701735; PubMed: 27672411]

Zalaudek 2006

* Zalaudek I, Argenziano G, Soyer HP, Corona R, Sera F, Blum A, et al. Three-point checklist of dermoscopy: an open internet study. British Journal of Dermatology 2006;154(3):431-7. [Other: ER4:15466171; PubMed: 16445771]

Excluded studies

Ahnlide 2013

Ahnlide I, Bjellerup M. Accuracy of clinical skin tumour diagnosis in a dermatological setting. Acta Dermato-Venereologica 2013;93(3):305-8.

Akasu 1996

Akasu R, Sugiyama H, Araki M, Ohtake N, Furue M, Tamaki K. Dermatoscopic and videomicroscopic features of melanocytic plantar nevi. American Journal of Dermatopathology 1996;18(1):10-8.

Alendar 2009

Alendar F, Drljevic I, Drljevic K, Alendar T. Early detection of melanoma skin cancer. Bosnian Journal of Basic Medical Sciences 2009;9(1):77-80.

Al Jalbout 2013

Al Jalbout S, Moscarella E, Longo C, Argenziano G, Piana S, Zalaudek I. Dermoscopy should always be performed... even in clear-cut cases! Journal of the American Academy of Dermatology 2013;69(4):e159-60.

Altamura 2006

Altamura D, Altobelli E, Micantonio T, Piccolo D, Fargnoli MC, Peris K. Dermoscopic patterns of acral melanocytic nevi and melanomas in a white population in central Italy. Archives of Dermatology 2006;142(9):1123-8.

Altamura 2010

Altamura D, Menzies SW, Argenziano G, Zalaudek I, Soyer HP, Sera F, et al. Dermatoscopy of basal cell carcinoma: morphologic variability of global and local features and accuracy of diagnosis. Journal of the American Academy of Dermatology 2010;62(1):67-75.

Amirnia 2016

Amirnia M, Ranjkesh MR, Azimpouran M, Karkon-Shayan F, Alikhah H, Jafari-Asl M, et al. Comparative Study of Dermatoscopic and Histopathologic Results in Facial Basal Cell Carcinoma and Melanocytic Nevi. Asian Pacific Journal of Cancer Prevention: Apjcp 2016;17(1):425-9.

Antonio 2013

Antonio JR, Soubhia RM, D'Avila SC, Caldas AC, Tridico LA, Alves FT. Correlation between dermoscopic and histopathological diagnoses of atypical nevi in a dermatology outpatient clinic of the Medical School of Sao Jose do Rio Preto, SP, Brazil. Anais Brasileiros de Dermatologia 2013;88(2):199-203.

Antoszewski 2015

Antoszewski B, Fijalkowska M, Stabryla P, Kasielska-Trojan A. Dermatoscopy as a Helpful Tool in Plastic Surgeon's Practice - A Preliminary Study. Polski Przeglad Chirurgiczny 2015;87(12):609-13.

Aoyagi 2010

Aoyagi S, Hata H, Izumi K, Iitani MM, Shimizu H. Diagnostic pitfalls of using dermoscopic features to differentiate between malignant melanoma and pigmented seborrhoeic keratosis. Acta Dermato-Venereologica 2010;90(4):440-1.

Argenziano 1997

Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Delfino M. Epiluminescence microscopy: criteria of cutaneous melanoma progression. Journal of the American Academy of Dermatology 1997;37(1):68-74.

Argenziano 1999

Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Delfino M. Clinical and dermatoscopic criteria for the preoperative evaluation of cutaneous melanoma thickness. Journal of the American Academy of Dermatology 1999;40(1):61-8.

Argenziano 2002

Argenziano G, Soyer HP, Chimenti S, Argenziano G, Ruocco V. Impact of dermoscopy on the clinical management of pigmented skin lesions. Clinics in Dermatology 2002;20(3):200-2.

Argenziano 2003

Argenziano G, Soyer HP, Chimenti S, Talamini R, Corona R, Sera F, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. Journal of the American Academy of Dermatology 2003;48(5):679-93.

Argenziano 2004

Argenziano G, Zalaudek I, Corona R, Sera F, Cicale L, Petrillo G, et al. Vascular structures in skin tumors: a dermoscopy study. Archives of Dermatology 2004;140(12):1485-9.

Argenziano 2004a

Argenziano G, Zalaudek I, Soyer HP. Which is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology? British Journal of Dermatology 2004;151(2):512-3.

Argenziano 2008

Argenziano G, Mordente I, Ferrara G, Sgambato A, Annese P, Zalaudek I. Dermoscopic monitoring of melanocytic skin lesions: clinical outcome and patient compliance vary according to follow-up protocols. British Journal of Dermatology 2008;159(2):331-6.

Argenziano 2010

Argenziano G, Kittler H, Ferrara G, Rubegni P, Malvehy J, Puig S, et al. Slow-growing melanoma: a dermoscopy follow-up study. British Journal of Dermatology 2010;162(2):267-73.

Argenziano 2011a

Argenziano G, Catricala C, Ardigo M, Buccini P, De Simone P, Eibenschutz L, et al. Dermoscopy of patients with multiple nevi: Improved management recommendations using a comparative diagnostic approach. Archives of Dermatology 2011;147(1):46-9.

Argenziano 2011b

Argenziano G, Longo C, Cameron A, Cavicchini S, Gourhant J Y, Lallas A, et al. Blue-black rule: a simple dermoscopic clue to recognize pigmented nodular melanoma. British Journal of Dermatology 2011;165(6):1251-5.

Argenziano 2012

Argenziano G, Zalaudek I, Hofmann-Wellenhof R, Bakos RM, Bergman W, Blum A, et al. Total body skin examination for skin cancer screening in patients with focused symptoms. Journal of the American Academy of Dermatology 2012;66(2):212-9.

Armstrong 2011

Armstrong A. Dermoscopy: An evidence-based approach for the early detection of melanoma. UNF Theses and Dissertations 2011;D.N.P.:65 p.

Ascierto 1998

Ascierto PA, Satriano RA, Palmieri G, Parasole R, Bosco L, Castello G. Epiluminescence microscopy as a useful approach in the early diagnosis of cutaneous malignant melanoma. Melanoma Research 1998;8(6):529-37.

Ascierto 2000

Ascierto PA, Palmieri G, Celentano E, Parasole R, Caraco C, Daponte A, et al. Sensitivity and specificity of epiluminescence microscopy: evaluation on a sample of 2731 excised cutaneous pigmented lesions. The Melanoma Cooperative Study. British Journal of Dermatology 2000;142(5):893-8.

Ascierto 2003

Ascierto PA, Palmieri G, Botti G, Satriano RA, Stanganelli I, Bono R, et al. Early diagnosis of malignant melanoma: Proposal of a working formulation for the management of cutaneous pigmented lesions from the Melanoma Cooperative Group. International Journal of Oncology 2003;22(6):1209-15.

Bafounta 2001

Bafounta ML, Beauchet A, Aegerter P, Saiag P. Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests. Archives of Dermatology 2001;137(10):1343-50.

Bajaj 2016

Bajaj S, Marchetti MA, Navarrete-Dechent C, Dusza SW, Kose K, Marghoob AA. The Role of Color and Morphologic Characteristics in Dermoscopic Diagnosis. JAMA Dermatology 2016;152(6):676-82.

Bauer 2005

Bauer J, Blum A, Strohhacker U, Garbe C. Surveillance of patients at high risk for cutaneous malignant melanoma using digital dermoscopy. British Journal of Dermatology 2005;152(1):87-92.

Bauer 2006

Bauer J, Leinweber B, Metzler G, Blum A, Hofmann-Wellenhof R, Leitz N, et al. Correlation with digital dermoscopic images can help dermatopathologists to diagnose equivocal skin tumours. British Journal of Dermatology 2006;155(3):546-51.

Benati 2015

Benati E, Argenziano G, Kyrgidis A, Moscarella E, Ciardo S, Bassoli S, et al. Melanoma and naevi with a globular pattern: confocal microscopy as an aid for diagnostic differentiation. British Journal of Dermatology 2015;173(5):1232-8.

Benelli 2000a

Benelli C, Roscetti E, Dal Pozzo V. Reproducibility of a dermoscopic method (7FFM) for the diagnosis of malignant melanoma. European Journal of Dermatology 2000;10(2):110-4.

Benvenuto-Andrade 2006

Benvenuto-Andrade C, Dusza SW, Hay JL, Agero AL, Halpern AC, Kopf AW, et al. Level of confidence in diagnosis: clinical examination versus dermoscopy examination. Dermatologic Surgery 2006;32(5):738-44.

Benvenuto-Andrade 2007

Benvenuto-Andrade C, Dusza SW, Agero AL, Scope A, Rajadhyaksha M, Halpern AC, et al. Differences between polarized light dermoscopy and immersion contact dermoscopy for the evaluation of skin lesions. Archives of Dermatology 2007;143(3):329-38.

Binder 1997

Binder M, Puespoeck-Schwarz M, Steiner A, Kittler H, Muellner M, Wolff K, et al. Epiluminescence microscopy of small pigmented skin lesions: short-term formal training improves the diagnostic performance of dermatologists. Journal of the American Academy of Dermatology 1997;36(2 Pt 1):197-202.

Blum 2003a

Blum A. Amelanotic/hypomelanotic melanoma--is dermatoscopy useful for diagnosis? Journal der Deutschen Dermatologischen Gesellschaft 2003;1(8):666-7.

Blum 2004a

Blum A. Pattern analysis, not simplified algorithms, is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology. British Journal of Dermatology 2004;151(2):511-2.

Blum 2004c

Blum A, Clemens J, Argenziano G. Three-colour test in dermoscopy: a re-evaluation. British Journal of Dermatology 2004;150(5):1040.

Blum 2004d

Blum A, Hofmann-Wellenhof R. Simplified dermoscopic diagnosis of acral melanocytic lesions: mountains and valleys. Australasian Journal of Dermatology 2004;45(4):235-6.

Blum 2006

Blum A, Clemens J, Argenziano G. Modified dermoscopic algorithm for the differentiation between melanocytic and nonmelanocytic skin tumors. Journal of Cutaneous Medicine & Surgery 2006;10(2):73-8.

Blum 2011

Blum A, Simionescu O, Argenziano G, Braun R, Cabo H, Eichhorn A, et al. Dermoscopy of pigmented lesions of the mucosa and the mucocutaneous junction: results of a multicenter study by the International Dermoscopy Society (IDS). Archives of Dermatology 2011;147(10):1181-7.

Blum 2014

Blum A, Ellwanger U, Luedtke H. Features Amplifying Dermoscopy (FAD) for better evaluation in difficult pigmented and non-pigmented melanocytic skin tumors. Journal der Deutschen Dermatologischen Gesellschaft 2014;12(1):77-9.

Boespflug 2015

Boespflug A, Guerra J, Dalle S, Thomas L. Enhancement of Customary Dermoscopy Education With Spaced Education e-Learning: A Prospective Controlled Trial. JAMA Dermatology 2015;151(8):847-53.

Bono 2001

Bono A, Maurichi A, Moglia D, Camerini T, Tragni G, Lualdi M, et al. Clinical and dermatoscopic diagnosis of early amelanotic melanoma. Melanoma Research 2001;11(5):491-4.

Borsari 2010

Borsari S, Longo C, Ferrari C, Benati E, Bassoli S, Schianchi S, et al. Dermoscopic island: a new descriptor for thin melanoma. Archives of Dermatology 2010;146(11):1257-62.

Bowns 2006

Bowns IR, Collins K, Walters SJ, McDonagh AJG. Telemedicine in dermatology: a randomised controlled trial. Health Technology Assessment (Winchester, England) 2006;10(43):iii-iv, ix-xi, 1-39.

Braun 2000

Braun RP, Krischer J, Saurat JH. The "wobble sign" in epiluminescence microscopy as a novel clue to the differential diagnosis of pigmented skin lesions. Archives of Dermatology 2000;136(7):940-2.

Braun 2007

Braun RP, Gaide O, Oliviero M, Kopf AW, French LE, Saurat JH, et al. The significance of multiple blue-grey dots (granularity) for the dermoscopic diagnosis of melanoma. British Journal of Dermatology 2007;157(5):907-13.

Braun-Falco 1990

Braun-Falco O, Stolz W, Bilek P, Merkle T, Landthaler M. The dermatoscope. A simplification of epiluminescent microscopy of pigmented skin changes. Hautarzt 1990;41(3):131-6.

Brown 2000

Brown N. Exploration of diagnostic techniques for malignant melanoma: an integrative review. Clinical Excellence for Nurse Practitioners 2000;4(5):263-71.

Buhl 2012

Buhl T, Hansen-Hagge C, Korpas B, Kaune KM, Haas E, Rosenberger A, et al. Integrating static and dynamic features of melanoma: the DynaMel algorithm. Journal of the American Academy of Dermatology 2012;66(1):27-36.

Bystryn 2003

Bystryn JC. Dermoscopic and histopathologic diagnosis of equivocal melanocytic skin lesions: an interdisciplinary study on 107 cases. Cancer 2003;97(7):1817; author reply 1817-8.

Cabrijan 2008

Cabrijan L, Lipozencic J, Batinac T, Lenkovic M, Gruber F, Stanic ZZ. Correlation between clinical-dermatoscopic and histopathologic diagnosis of skin tumors in our patients. Collegium Antropologicum 2008;32 Suppl 2:195-7.

Canpolat 2011

Canpolat F, Akış HK, Akay BN, Erdem C. Dermoscopic Features of Acral Melanocytic Nevi [Akral Melanositik Nevüslerin Dermoskopik Özellikleri]. Archives of the Turkish Dermatology & Venerology / Turkderm 2011;45(4):193-7.

Cardenas 2009

Cardenas E, Sosa A, Bezaury P, La Madrid JV, Reyes E, Topete RO. Usefulness of high resolution ultrasound of 17 Mhz in palpable skin lesions. An analysis of 27 patients [Utilidad del ultrasonido de alta resolucion de 17 MHz en lesiones cutaneas palpables. Analisis de 27 pacientes]. Dermatologia Revista Mexicana 2009;53(3):119-24.

Carli 1998

Carli P, De Giorgi V, Naldi L, Dosi G. Reliability and inter-observer agreement of dermoscopic diagnosis of melanoma and melanocytic naevi. Dermoscopy Panel. European Journal of Cancer Prevention 1998;7(5):397-402.

Carli 2000

Carli P, De Giorgi V, Massi D, Giannotti B. The role of pattern analysis and the ABCD rule of dermoscopy in the detection of histological atypia in melanocytic naevi. British Journal of Dermatology 2000;143(2):290-7.

Carli 2003a

Carli P, Mannone F, De Giorgi V, Nardini P, Chiarugi A, Giannotti B. The problem of false-positive diagnosis in melanoma screening: the impact of dermoscopy. Melanoma Research 2003;13(2):179-82.

Carli 2004

Carli P, de Giorgi V, Chiarugi A, Nardini P, Weinstock MA, Crocetti E, et al. Addition of dermoscopy to conventional naked-eye examination in melanoma screening: a randomized study. Journal of the American Academy of Dermatology 2004;50(5):683-9.

Carli 2004a

Carli P, De Giorgi V, Crocetti E, Mannone F, Massi D, Chiarugi A, et al. Improvement of malignant/benign ratio in excised melanocytic lesions in the 'dermoscopy era': a retrospective study 1997-2001. British Journal of Dermatology 2004;150(4):687-92.

Carli 2005

Carli P, Chiarugi A, De Giorgi V. Examination of lesions (including dermoscopy) without contact with the patient is associated with improper management in about 30% of equivocal melanomas. Dermatologic Surgery 2005;31(2):169-72.

Carlos-Ortega 2007

Carlos-Ortega B, Sanchez-Alva ME, Ysita-Morales A, Angeles-Garay U. Correlation among simple observation and dermoscopy in the study of pigmented lesions of the skin. Revista Medica del Instituto Mexicano del Seguro Social 2007;45(6):541-8.

Carroll 1998

Carroll DM, Billingsley EM, Helm KF. Diagnosing basal cell carcinoma by dermatoscopy. Journal of Cutaneous Medicine & Surgery 1998;3(2):62-7.

Chen 2013

Chen LL, Liebman TN, Soriano RP, Dusza SW, Halpern AC, Marghoob AA. One-year follow-up of dermoscopy education on the ability of medical students to detect skin cancer. Dermatology 2013;226(3):267-73.

Ciudad-Blanco 2014

Ciudad-Blanco C, Aviles-Izquierdo JA, Lazaro-Ochaita P, Suarez-Fernandez R. Dermoscopic findings for the early detection of melanoma: an analysis of 200 cases. Actas Dermo-Sifiliograficas 2014;105(7):683-93.

de Giorgi 2006

de Giorgi V, Trez E, Salvini C, Duquia R, De Villa D, Sestini S, et al. Dermoscopy in black people. British Journal of Dermatology 2006;155(4):695-9.

De Giorgi 2011

* De Giorgi V, Grazzini M, Rossari S, Gori A, Alfaioli B, Papi F, et al. Adding dermatoscopy to naked eye examination of equivocal melanocytic skin lesions: effect on intention to excise by general dermatologists. Clinical & Experimental Dermatology 2011;36:255-9. [Other: ER4:15465901]

Delfino 1997

Delfino M, Fabbrocini G, Argenziano G, Magliocchetti N, Nofroni I. A statistical analysis of the characteristics of pigmented skin lesions using epiluminescence microscopy. Journal of the European Academy of Dermatology and Venereology 1997;9(3):243-8.

de Troya-Martin 2008

de Troya-Martin M, Blazquez-Sanchez N, Fernandez-Canedo I, Frieyro-Elicegui M, Funez-Liebana R, Rivas-Ruiz F. [Dermoscopic study of cutaneous malignant melanoma: descriptive analysis of 45 cases]. Actas Dermo-Sifiliograficas 2008;99(1):44-53.

Di Chiacchio 2010

Di Chiacchio N, Hirata SH, Enokihara MY, Michalany NS, Fabbrocini G, Tosti A. Dermatologists' accuracy in early diagnosis of melanoma of the nail matrix. Archives of Dermatology 2010;146(4):382-7.

Di Stefani 2007

Di Stefani A, Zalaudek I, Argenziano G, Chimenti S, Soyer HP. Feasibility of a two-step teledermatologic approach for the management of patients with multiple pigmented skin lesions. Dermatologic Surgery 2007;33(6):686-92.

Dummer 1995

Dummer W, Blaheta HJ, Bastian BC, Schenk T, Brocker EV, Remy W. Preoperative characterization of pigmented skin lesions by epiluminescence microscopy and high-frequency ultrasound. Archives of Dermatology 1995;131(3):279-85.

Elwan 2016

Elwan NM, Eltatawy RA, Elfar NN, Elsakka OM. Dermoscopic features of acral pigmented lesions in Egyptian patients: a descriptive study. International Journal of Dermatology 2016;55(2):187-92.

Fabbrocini 2008

Fabbrocini G, Balato A, Rescigno O, Mariano M, Scalvenzi M, Brunetti B. Telediagnosis and face-to-face diagnosis reliability for melanocytic and non-melanocytic 'pink' lesions. Journal of the European Academy of Dermatology & Venereology 2008;22(2):229-34.

Ferrara 2002

Ferrara G, Argenziano G, Soyer HP, Corona R, Sera F, Brunetti B, et al. Dermoscopic and histopathologic diagnosis of equivocal melanocytic skin lesions: an interdisciplinary study on 107 cases. Cancer 2002;95(5):1094-100.

Fidalgo 2003

Fidalgo A, Caldas Lopes L, Macedo Ferreira A. Digital dermatoscopy: One-year experience with the DANAOS system. Skin Cancer 2003;18(4):211-8.

Fruhauf 2012

Fruhauf J, Leinweber B, Fink-Puches R, Ahlgrimm-Siess V, Richtig E, Wolf I H, et al. Patient acceptance and diagnostic utility of automated digital image analysis of pigmented skin lesions. Journal of the European Academy of Dermatology & Venereology 2012;26(3):368-72.

Fueyo-Casado 2009

Fueyo-Casado A, Vazquez-Lopez F, Sanchez-Martin J, Garcia-Garcia B, Perez-Oliva N. Evaluation of a program for the automatic dermoscopic diagnosis of melanoma in a general dermatology setting. Dermatologic Surgery 2009;35(2):257-9; discussion 260-2.

Giacomel 2005

Giacomel J, Zalaudek I. Dermoscopy of superficial basal cell carcinoma. Dermatologic Surgery 2005;31(12):1710-3.

Giacomel 2014

Giacomel J, Lallas A, Zalaudek I, Argenziano G. Dermoscopic "signature" pattern of pigmented and nonpigmented lentigo maligna. Journal of the American Academy of Dermatology 2014;70(2):e33-5.

Giannotti 2004

Giannotti B, Carli P. Improvement of early diagnosis of melanoma in a mediterranean population: The experience of the Florence melanoma clinic. [Italian, English] Novita in tema di diagnosi precoce del melanoma cutaneo: L'esperienza del gruppo Fiorentino. Giornale Italiano di Dermatologia e Venereologia 2004;139(2):89-96.

Gill 2015

Gill L, Wang S, Mancebo SE, Lim HW, Kohen LL. Dermoscopic features of acral melanocytic nevi in patients with skin types V and VI: A cross-sectional study. Journal of the American Academy of Dermatology 2015;73(6):1059-61.

Gilmore 2009

Gilmore S, Hofmann-Wellenhof R, Muir J, Soyer HP. Lacunarity analysis: a promising method for the automated assessment of melanocytic naevi and melanoma. PLoS ONE [Electronic Resource] 2009;4(10):e7449.

Grichnik 2003

Grichnik JM. Dermoscopy of melanocytic neoplasms: subpatterns of dysplastic/atypical nevi. Archives of Dermatology 2003;139(12):1696.

Grichnik 2004

Grichnik JM. Dermoscopy of melanocytic neoplasms: familial patterns. Archives of Dermatology 2004;140(5):642.

Guillod 1996

Guillod JF, Schmid Ph, Fischer S, Salomon D, Saurat JH. Detection and classification of pigmented skin lesions by dermatoscopic digital image processing. Dermatology 1996;193(2):169.

Gunduz 2003

Gunduz K, Koltan S, Sahin MT, E Filiz E. Analysis of melanocytic naevi by dermoscopy during pregnancy. Journal of the European Academy of Dermatology & Venereology 2003;17(3):349-51.

Hacioglu 2013

Hacioglu S, Saricaoglu H, Baskan EB, Uner SI, Aydogan K, Tunali S. The value of spectrophotometric intracutaneous analysis in the noninvasive diagnosis of nonmelanoma skin cancers. Clinical & Experimental Dermatology 2013;38(5):464-9.

Haenssle 2006

Haenssle HA, Krueger U, Vente C, Thoms KM, Bertsch HP, Zutt M, et al. Results from an observational trial: digital epiluminescence microscopy follow-up of atypical nevi increases the sensitivity and the chance of success of conventional dermoscopy in detecting melanoma. Journal of Investigative Dermatology 2006;126(5):980-5.

Haenssle 2010

Haenssle HA, Korpas B, Hansen-Hagge C, Buhl T, Kaune KM, Johnsen S, et al. Selection of patients for long-term surveillance with digital dermoscopy by assessment of melanoma risk factors. Archives of Dermatology 2010;146(3):257-64.

Haspeslagh 2016

Haspeslagh M, Vossaert K, Lanssens S, Noe M, Hoorens I, Chevolet I, et al. Comparison of Ex Vivo and In Vivo Dermoscopy in Dermatopathologic Evaluation of Skin Tumors. JAMA Dermatology 2016;152(3):312-7.

Henning 2007

Henning JS, Dusza SW, Wang SQ, Marghoob AA, Rabinovitz HS, Polsky D, et al. The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy. Journal of the American Academy of Dermatology 2007;56(1):45-52.

Henning 2008

Henning JS, Stein JA, Yeung J, et al.. CASH algorithm for dermoscopy revisited. Archives of Dermatology 2008;144:554-555.

Herschorn 2012

Herschorn A. Dermoscopy for melanoma detection in family practice. Canadian Family Physician 2012;58(7):740-5, e372-8.

Hirata 2011

Hirata SH, Yamada S, Enokihara MY, Di Chiacchio N, de Almeida FA, Enokihara MM, et al. Patterns of nail matrix and bed of longitudinal melanonychia by intraoperative dermatoscopy. Journal of the American Academy of Dermatology 2011;65(2):297-303.

Hoffmann 2003

Hoffmann K, Gambichler T, Rick A, Kreutz M, Anschuetz M, Grunendick T, et al. Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy. British Journal of Dermatology 2003;149(4):801-9.

Hoorens 2016

Hoorens I, Vossaert K, Pil L, Boone B, De Schepper S, Ongenae K, et al. Total-Body Examination vs Lesion-Directed Skin Cancer Screening. JAMA Dermatology 2016;152(1):27-34.

Ishioka 2009

Ishioka P, Tenorio JM, Lopes PRI, Yamada S, Michalany NS, Amaral MB, et al. A comparative study of teledermatoscopy and face-to-face examination of pigmented skin lesions. Journal of Telemedicine & Telecare 2009;15(5):221-5.

Iyatomi 2006

Iyatomi H, Oka H, Saito M, Miyake A, Kimoto M, Yamagami J, et al. Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system. Melanoma Research 2006;16(2):183-90.

Iyatomi 2008

Iyatomi H, Oka H, Celebi ME, Ogawa K, Argenziano G, Soyer HP, et al. Computer-based classification of dermoscopy images of melanocytic lesions on acral volar skin. Journal of Investigative Dermatology 2008;128(8):2049-54.

Johr 2002

Johr RH. Dermoscopy: alternative melanocytic algorithms-the ABCD rule of dermatoscopy, Menzies scoring method, and 7-point checklist. Clinics in Dermatology 2002;20(3):240-7.

Kawabata 1998

Kawabata Y, Tamaki K. Distinctive dermatoscopic features of acral lentiginous melanoma in situ from plantar melanocytic nevi and their histopathologic correlation. Journal of Cutaneous Medicine & Surgery 1998;2(4):199-204.

Kawabata 2001

Kawabata Y, Ohara K, Hino H, Tamaki K. Two kinds of Hutchinson's sign, benign and malignant. Journal of the American Academy of Dermatology 2001;44(2, Part 1):305-7.

Kefel 2012

Kefel S, Guvenc P, LeAnder R, Stricklin SM, Stoecker WV. Discrimination of basal cell carcinoma from benign lesions based on extraction of ulcer features in polarized-light dermoscopy images. Skin Research & Technology 2012;18(4):471-5.

Kenet 1994

Kenet RO, Fitzpatrick TB. Reducing mortality and morbidity of cutaneous melanoma: a six year plan. B). Identifying high and low risk pigmented lesions using epiluminescence microscopy. Journal of Dermatology 1994;21(11):881-4.

Kittler 2002

Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. Lancet Oncology 2002;3(3):159-65.

Kittler 2006

Kittler H. Value of follow-up of pigmented skin lesions by digital dermatoscopy. Journal of Investigative Dermatology 2006;126:S20.

Koga 2011

Koga H, Saida T. Revised 3-step dermoscopic algorithm for the management of acral melanocytic lesions. Archives of Dermatology 2011;147(6):741-3.

Korotkov 2012

Korotkov K, Garcia R. Computerized analysis of pigmented skin lesions: a review. Artificial Intelligence in Medicine 2012;56(2):69-90.

Lallas 2015

Lallas A, Kyrgidis A, Koga H, Moscarella E, Tschandl P, Apalla Z, et al. The BRAAFF checklist: a new dermoscopic algorithm for diagnosing acral melanoma. British Journal of Dermatology 2015;173(4):1041-9.

Liebman 2011

Liebman TN, Scope A, Rabinovitz H, Braun RP, Marghoob AA. Rosettes may be observed in a range of conditions. Archives of Dermatology 2011;147(12):1468.

Liebman 2012

Liebman TN, Rabinovitz HS, Balagula Y, Jaimes-Lopez N, Marghoob AA. White shiny structures in melanoma and BCC. Archives of Dermatology 2012;148(1):146.

Lipoff 2008

Lipoff JB, Scope A, Dusza SW, Marghoob AA, Oliveria SA, Halpern AC. Complex dermoscopic pattern: a potential risk marker for melanoma. British Journal of Dermatology 2008;158(4):821-4.

Liu 2012

Liu Z, Sun J, Smith L, Smith M, Warr R. Distribution quantification on dermoscopy images for computer-assisted diagnosis of cutaneous melanomas. Medical & Biological Engineering & Computing 2012;50(5):503-13.

Lorentzen 1999a

Lorentzen H, Weismann K, Secher L, Petersen CS, Larsen FG. The dermatoscopic ABCD rule does not improve diagnostic accuracy of malignant melanoma. Acta Dermato-Venereologica 1999;79(6):469-72.

Luttrell 2012

Luttrell MJ, McClenahan P, Hofmann-Wellenhof R, Fink-Puches R, Soyer HP. Laypersons' sensitivity for melanoma identification is higher with dermoscopy images than clinical photographs. British Journal of Dermatology 2012;167(5):1037-41.

MacKie 1971

MacKie RM. An aid to the preoperative assessment of pigmented lesions of the skin. British Journal of Dermatology 1971;85(3):232-8.

MacKie 2002

Mackie RM, Fleming C, McMahon AD, Jarrett P. The use of the dermatoscope to identify early melanoma using the three-colour test. British Journal of Dermatology 2002;146(3):481-4.

Markowitz 2015

Markowitz O, Schwartz M, Feldman E, Bienenfeld A, Bieber AK, Ellis J, et al. Evaluation of Optical Coherence Tomography as a Means of Identifying Earlier Stage Basal Cell Carcinomas while Reducing the Use of Diagnostic Biopsy. The Journal of Clinical & Aesthetic Dermatology 2015;8(10):14-20.

Massi 2001

Massi D, De Giorgi V, Carli P, Santucci M. Diagnostic significance of the blue hue in dermoscopy of melanocytic lesions: a dermoscopic-pathologic study. American Journal of Dermatopathology 2001;23(5):463-9.

Mayer 1997

Mayer J. Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma. Medical Journal of Australia 1997;167(4):206-10.

Menzies 1996a

Menzies SW, Ingvar C, McCarthy WH. A sensitivity and specificity analysis of the surface microscopy features of invasive melanoma. Melanoma Research 1996;6(1):55-62.

Menzies 1999

Menzies SW. Automated epiluminescence microscopy: human vs machine in the diagnosis of melanoma. Archives of Dermatology 1999;135(12):1538-40.

Menzies 2000

Menzies SW, Westerhoff K, Rabinovitz H, Kopf AW, McCarthy WH, Katz B. Surface microscopy of pigmented basal cell carcinoma. Archives of Dermatology 2000;136(8):1012-6.

Menzies 2001

Menzies SW, Gutenev A, Avramidis M, Batrac A, McCarthy WH. Short-term digital surface microscopic monitoring of atypical or changing melanocytic lesions. Archives of Dermatology 2001;137(12):1583-9.

Mun 2016

Mun JH, Ohn J, Kim WI, Park SM, Kim MB. Dermoscopy of Melanomas on the Trunk and Extremities in Asians. PLoS ONE [Electronic Resource] 2016;11(7):e0158374.

Nathansohn 2007

Nathansohn N, Orenstein A, Trau H, Liran A, Schachter J. Pigmented lesions clinic for early detection of melanoma: preliminary results. Israel Medical Association Journal: Imaj 2007;9(10):708-12.

Navarrete-Dechent 2016

Navarrete-Dechent C, Bajaj S, Marchetti MA, Rabinovitz H, Dusza SW, Marghoob AA. Association of Shiny White Blotches and Strands With Nonpigmented Basal Cell Carcinoma: Evaluation of an Additional Dermoscopic Diagnostic Criterion. JAMA Dermatology 2016;152(5):546-52.

Pan 2008

Pan Y, Chamberlain AJ, Bailey M, Chong AH, Haskett M, Kelly JW. Dermatoscopy aids in the diagnosis of the solitary red scaly patch or plaque-features distinguishing superficial basal cell carcinoma, intraepidermal carcinoma, and psoriasis. Journal of the American Academy of Dermatology 2008;59(2):268-74.

Panasiti 2009

Panasiti V, Devirgiliis V, Curzio M, Roberti V, Gobbi S, Masciangelo R, et al. The reticular point of view in dermatoscopy. Journal of the American Academy of Dermatology 2009;61(4):605-10.

Pazzini 1996

Pazzini C, Pozzi M, Betti R, Vergani R, Crosti C. Improvement of diagnostic accuracy in the clinical diagnosis of pigmented skin lesions by epiluminescence microscopy. Skin Cancer 1996;11(2):159-61.

Pehamberger 1987

Pehamberger H, Steiner A, Wolff K. In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions. Journal of the American Academy of Dermatology 1987;17(4):571-83.

Pellacani 2002

Pellacani G, Seidenari S. Comparison between morphological parameters in pigmented skin lesion images acquired by means of epiluminescence surface microscopy and polarized-light videomicroscopy. Clinics in Dermatology 2002;20(3):222-7.

Pellacani 2006

Pellacani G, Grana C, Seidenari S. Algorithmic reproduction of asymmetry and border cut-off parameters according to the ABCD rule for dermoscopy. Journal of the European Academy of Dermatology & Venereology 2006;20(10):1214-9.

Pellacani 2007

Pellacani G, Bassoli S, Longo C, Cesinaro AM, Seidenari S. Diving into the blue: in vivo microscopic characterization of the dermoscopic blue hue. Journal of the American Academy of Dermatology 2007;57(1):96-104.

Pellacani 2009

Pellacani G, Longo C, Ferrara G, Cesinaro AM, Bassoli S, Guitera P, et al. Spitz nevi: In vivo confocal microscopic features, dermatoscopic aspects, histopathologic correlates, and diagnostic significance. Journal of the American Academy of Dermatology 2009;60(2):236-47.

Peris 2002

Peris K, Altobelli E, Ferrari A, Fargnoli MC, Piccolo D, Esposito M, et al. Interobserver agreement on dermoscopic features of pigmented basal cell carcinoma. Dermatologic Surgery 2002;28(7):643-5.

Peris 2002a

Peris K, Ferrari A, Argenziano G, Soyer HP, Chimenti S. Dermoscopic classification of Spitz/Reed nevi. Clinics in Dermatology 2002;20(3):259-62.

Phan 2010

Phan A, Dalle S, Touzet S, Ronger-Savle S, Balme B, Thomas L. Dermoscopic features of acral lentiginous melanoma in a large series of 110 cases in a white population. British Journal of Dermatology 2010;162(4):765-71.

Piccolo 2002a

Piccolo D, Peris K, Chimenti S, Argenziano G, Soyer HP. Jumping into the future using teledermoscopy. SKINmed 2002;1(1):20-4.

Piccolo 2004

Piccolo D, Soyer HP, Chimenti S, Argenziano G, Bartenjev I, Hofmann-Wellenhof R, et al. Diagnosis and categorization of acral melanocytic lesions using teledermoscopy. Journal of Telemedicine & Telecare 2004;10(6):346-50.

Piccolo 2006

Piccolo D, Fargnoli MC, Ferrara G, Lozzi GP, Altamura D, Ventura T, et al. Hypoepiluminescence microscopy of pigmented skin lesions: new approach to improve recognition of dermoscopic structures. Dermatologic Surgery 2006;32(11):1391-7.

Pizzichetta 2001

Pizzichetta MA, Argenziano G, Talamini R, Piccolo D, Gatti A, Trevisan G, et al. Dermoscopic criteria for melanoma in situ are similar to those for early invasive melanoma. Cancer 2001;91(5):992-7.

Pizzichetta 2001a

Pizzichetta MA, Talamini R, Piccolo D, Argenziano G, Pagnanelli G, Burgdorf T, et al. The ABCD rule of dermatoscopy does not apply to small melanocytic skin lesions. Archives of Dermatology 2001;137(10):1376-8.

Pizzichetta 2007

Pizzichetta MA, Stanganelli I, Bono R, Soyer HP, Magi S, Canzonieri V, et al. Dermoscopic features of difficult melanoma. Dermatologic Surgery 2007;33(1):91-9.

Pizzichetta 2010

Pizzichetta MA, Canzonieri V, Massarut S, Baresic T, Borsatti E, Menzies SW. Pitfalls in the dermoscopic diagnosis of amelanotic melanoma. Journal of the American Academy of Dermatology 2010;62(5):893-4.

Pizzichetta 2013

Pizzichetta MA, Talamini R, Marghoob AA, Soyer HP, Argenziano G, Bono R, et al. Negative pigment network: an additional dermoscopic feature for the diagnosis of melanoma. Journal of the American Academy of Dermatology 2013;68(4):552-9.

Pralong 2012

Pralong P, Bathelier E, Dalle S, Poulalhon N, Debarbieux S, Thomas L. Dermoscopy of lentigo maligna melanoma: report of 125 cases. British Journal of Dermatology 2012;167(2):280-7.

Provost 1998

Provost N, Kopf AW, Rabinovitz HS, Stolz W, DeDavid M, Wasti Q, et al. Comparison of conventional photographs and telephonically transmitted compressed digitized images of melanomas and dysplastic nevi. Dermatology 1998;196(3):299-304.

Rader 2014

Rader RK, Payne KS, Guntupalli U, Rabinovitz HS, Oliviero MC, Drugge RJ, et al. The pink rim sign: location of pink as an indicator of melanoma in dermoscopic images. Journal of Skin Cancer 2014;2014:719740.

Rajpara 2009

Rajpara SM, Botello AP, Townend J, Ormerod AD. Systematic review of dermoscopy and digital dermoscopy/ artificial intelligence for the diagnosis of melanoma. British Journal of Dermatology 2009;161(3):591-604.

Reggiani 2015

Reggiani C, Manfredini M, Mandel VD, Farnetani F, Ciardo S, Bassoli S, et al. Update on non-invasive imaging techniques in early diagnosis of non-melanoma skin cancer. Giornale Italiano di Dermatologia e Venereologia 2015;150(4):393-405.

Rigel 1997

Rigel DS. Epiluminescence microscopy in clinical diagnosis of pigmented skin lesions? Lancet 1997;349(9065):1566-7.

Ronger 2002

Ronger S, Touzet S, Ligeron C, Balme B, Viallard AM, Barrut D, et al. Dermoscopic examination of nail pigmentation. Archives of Dermatology 2002;138(10):1327-33.

Rosendahl 2012

Rosendahl C, Cameron A, Argenziano G, Zalaudek I, Tschandl P, Kittler H. Dermoscopy of squamous cell carcinoma and keratoacanthoma. Archives of Dermatology 2012;148(12):1386-92.

Rosendahl 2012a

Rosendahl C, Cameron A, McColl I, Wilkinson D. Dermatoscopy in routine practice: 'Chaos and Clues'. Australian Family Physician 2012;41(7):482-7.

Rossi 2000

Rossi CR, Vecchiato A, Bezze G, Mastrangelo G, Montesco MC, Mocellin S, et al. Early detection of melanoma: an educational campaign in Padova, Italy. Melanoma Research 2000;10(2):181-7.

Rubegni 2002

Rubegni P, Burroni M, Dell'eva G, Andreassi L. Digital dermoscopy analysis for automated diagnosis of pigmented skin lesions. Clinics in Dermatology 2002;20(3):309-12.

Rubegni 2005

Rubegni P, Burroni M, Andreassi A, Fimiani M. The role of dermoscopy and digital dermoscopy analysis in the diagnosis of pigmented skin lesions. Archives of Dermatology 2005;141(11):1444-6.

Rubegni 2005a

Rubegni P, Burroni M, Sbano P, Andreassi L. Digital dermoscopy analysis and internet-based program for discrimination of pigmented skin lesion dermoscopic images. British Journal of Dermatology 2005;152(2):395-6.

Rubegni 2010

Rubegni P, Cevenini G, Burroni M, Bono R, Sbano P, Biagioli M, et al. Objective follow-up of atypical melanocytic skin lesions: a retrospective study. Archives of Dermatological Research 2010;302(7):551-60.

Sahin 2004

Sahin MT, Ozturkcan S, Ermertcan AT, Gunes AT. A comparison of dermoscopic features among lentigo senilis/initial seborrheic keratosis, seborrheic keratosis, lentigo maligna and lentigo maligna melanoma on the face. Journal of Dermatology 2004;31(11):884-9.

Saida 2002

Saida T, Oguchi S, Miyazaki A. Dermoscopy for acral pigmented skin lesions. Clinics in Dermatology 2002;20(3):279-85.

Saida 2004

Saida T, Miyazaki A, Oguchi S, Ishihara Y, Yamazaki Y, Murase S, et al. Significance of dermoscopic patterns in detecting malignant melanoma on acral volar skin: results of a multicenter study in Japan. Archives of Dermatology 2004;140(10):1233-8.

Sakakibara 2010

Sakakibara A, Kamijima M, Shibata S, Yasue S, Kono M, Tomita Y. Dermoscopic evaluation of vascular structures of various skin tumors in Japanese patients. Journal of Dermatology 2010;37(4):316-22.

Salerni 2011

Salerni G, Lovatto L, Carrera C, Palou J, Alos L, Puig-Butille JA, et al. Correlation among dermoscopy, confocal reflectance microscopy, and histologic features of melanoma and basal cell carcinoma collision tumor. Dermatologic Surgery 2011;37(2):275-9.

Salerni 2012

Salerni G, Carrera C, Lovatto L, Puig-Butille JA, Badenas C, Plana E, et al. Benefits of total body photography and digital dermatoscopy ("two-step method of digital follow-up") in the early diagnosis of melanoma in patients at high risk for melanoma. Journal of the American Academy of Dermatology 2012;67(1):e17-27.

Salerni 2013

Salerni G, Teran T, Puig S, Malvehy J, Zalaudek I, Argenziano G, et al. Meta-analysis of digital dermoscopy follow-up of melanocytic skin lesions: a study on behalf of the International Dermoscopy Society. Journal of the European Academy of Dermatology & Venereology 2013;27(7):805-14.

Salvio 2011

Salvio AG, Assumpcao JA, Segalla JGM, Panfilo BL, Nicolini HR, Didone R. One year experience of a model for melanoma continuous prevention in the city of Jau (Sao Paulo), Brazil. Anais Brasileiros de Dermatologia 2011;86(4):669-74.

Sanchez-Martin 2012

Sanchez-Martin J, Vazquez-Lopez F, Perez-Oliva N, Argenziano G. Dermoscopy of small basal cell carcinoma: study of 100 lesions 5 mm or less in diameter. Dermatologic Surgery 2012;38(6):947-50.

Savk 2004

Savk E, Sahinkaras E, Okyay P, Karaman G, Erkek M, Sendur N. Interobserver agreement in the use of the ABCD rule for dermoscopy. Journal of Dermatology 2004;31(12):1041-3.

Sawada 2013

Sawada M, Tanaka M. Self-assembly of a simple low-cost dermoscope for examination of skin lesions. Dermatology Practical & Conceptual 2013;3(4):35.

Sboner 2003

Sboner A, Eccher C, Blanzieri E, Bauer P, Cristofolini M, Zumiani G, et al. A multiple classifier system for early melanoma diagnosis. Artificial Intelligence in Medicine 2003;27(1):29-44.

Schulz 2001

Schulz H. [Epiluminescent microscopy aspects of initial cutaneous melanoma metastases]. Hautarzt 2001;52(1):21-5.

Scope 2015

Scope A, Braun RP. The Recognition Process in Dermoscopy: Analytic Approach vs Heuristic Approach. JAMA Dermatology 2015;151(7):704-6.

Segura 2009

Segura S, Puig S, Carrera C, Palou J, Malvehy J. Development of a two-step method for the diagnosis of melanoma by reflectance confocal microscopy. Journal of the American Academy of Dermatology 2009;61(2):216-29.

Seidenari 2004

Seidenari S, Pellacani G, Righi E, Di Nardo A. Is JPEG compression of videomicroscopic images compatible with telediagnosis? Comparison between diagnostic performance and pattern recognition on uncompressed TIFF images and JPEG compressed ones. Telemedicine Journal & E-Health 2004;10(3):294-303.

Seidenari 2006

Seidenari S, Longo C, Giusti F, Pellacani G. Clinical selection of melanocytic lesions for dermoscopy decreases the identification of suspicious lesions in comparison with dermoscopy without clinical preselection. British Journal of Dermatology 2006;154(5):873-9.

Seidenari 2006a

Seidenari S, Pellacani G, Grana C. Asymmetry in dermoscopic melanocytic lesion images: a computer description based on colour distribution. Acta Dermato-Venereologica 2006;86(2):123-8.

Seidenari 2012

Seidenari S, Ferrari C, Borsari S, Bassoli S, Cesinaro AM, Giusti F, et al. The dermoscopic variability of pigment network in melanoma in situ. Melanoma Research 2012;22(2):151-7.

Seidenari 2013

Seidenari S, Arginelli F, Dunsby C, French PMW, Konig K, Magnoni C, et al. Multiphoton laser tomography and fluorescence lifetime imaging of melanoma: morphologic features and quantitative data for sensitive and specific non-invasive diagnostics. PLoS ONE [Electronic Resource] 2013;8(7):e70682.

Serrao 2006

Serrao VV, Baptista J, Paris F, Lopes LC, Fidalgo A, Ferreira A. Digital dermoscopy. Review of 652 lesions analysed by the DANAOS system. Skin Cancer 2006;21(4):185-98.

Sgouros 2014

Sgouros D, Lallas A, Julian Y, Rigopoulos D, Zalaudek I, Longo C, et al. Assessment of SIAscopy in the triage of suspicious skin tumours. Skin Research & Technology 2014;20(4):440-4.

Shakya 2012

Shakya NM, LeAnder RW, Hinton KA, Stricklin SM, Rader RK, Hagerty J, et al. Discrimination of squamous cell carcinoma in situ from seborrheic keratosis by color analysis techniques requires information from scale, scale-crust and surrounding areas in dermoscopy images. Computers in Biology & Medicine 2012;42(12):1165-9.

Shitara 2014

Shitara D, Ishioka P, Alonso-Pinedo Y, Palacios-Bejarano L, Carrera C, Malvehy J, et al. Shiny white streaks: a sign of malignancy at dermoscopy of pigmented skin lesions. Acta Dermato-Venereologica 2014;94(2):132-7.

Shitara 2015

Shitara D, Nascimento M, Ishioka P, Carrera C, Alos L, Malvehy J, et al. Dermoscopy of Naevus-associated Melanomas. Acta Dermato-Venereologica 2015;95(6):671-5.

Sondak 2015

Sondak VK, Glass LF, Geller AC. Risk-stratified screening for detection of melanoma. JAMA 2015;313(6):616-7.

Soyer 1987

Soyer HP, Smolle J, Kerl H, Stettner H. Early diagnosis of malignant melanoma by surface microscopy. Lancet 1987;2(8562):803.

Soyer 2001

Soyer HP, Argenziano G, Talamini R, Chimenti S. Is dermoscopy useful for the diagnosis of melanoma? Archives of Dermatology 2001;137(10):1361-3.

Stanganelli 1998a

Stanganelli I, Bucchi L. Epiluminescence microscopy versus clinical evaluation of pigmented skin lesions: effects of Operator's training on reproducibility and accuracy. Dermatology and Venereology Society of the Canton of Ticino. Dermatology 1998;196(2):199-203.

Steiner 1987

* Steiner A, Pehamberger H, Wolff K. In vivo epiluminescence microscopy of pigmented skin lesions. II. Diagnosis of small pigmented skin lesions and early detection of malignant melanoma. Journal of the American Academy of Dermatology 1987;17(4):584-91. [Other: ER4:17940992; PubMed: 3668003]

Steiner 1987a

Steiner A, Pehamberger H, Wolff K. Improvement of the diagnostic accuracy in pigmented skin lesions by epiluminescent light microscopy. Anticancer research 1987;7(3 Pt B):433-4.

Steiner 1993

Steiner A, Binder M, Schemper M, Wolff K, Pehamberger H. Statistical evaluation of epiluminescence microscopy criteria for melanocytic pigmented skin lesions. Journal of the American Academy of Dermatology 1993;29(4):581-8.

Stephens 2013

Stephens A, Fraga-Braghiroli N, Oliviero M, Rabinovitz H, Scope A. Spoke wheel-like structures in superficial basal cell carcinoma: a correlation between dermoscopy, histopathology, and reflective confocal microscopy. Journal of the American Academy of Dermatology 2013;69(5):e219-21.

Stoecker 2009

Stoecker WV, Gupta K, Shrestha B, Wronkiewiecz M, Chowdhury R, Stanley RJ, et al. Detection of basal cell carcinoma using color and histogram measures of semitranslucent areas. Skin Research & Technology 2009;15(3):283-7.

Stoecker 2009a

Stoecker WV, Kolm I, Rabinovitz HS, Oliviero MC, Xu J, Malters JM. Semitranslucency in dermoscopic images of basal cell carcinoma. Archives of Dermatology 2009;145(2):224.

Stoecker 2011

Stoecker WV, Wronkiewiecz M, Chowdhury R, Stanley RJ, Xu J, Bangert A, et al. Detection of granularity in dermoscopy images of malignant melanoma using color and texture features. Computerized Medical Imaging & Graphics 2011;35(2):144-7.

Stolz 2002

Stolz W, Schiffner R, Burgdorf WHC. Dermatoscopy for facial pigmented skin lesions. Clinics in Dermatology 2002;20(3):276-8.

Stratigos 2007

Stratigos A, Nikolaou V, Kedicoglou S, Antoniou C, Stefanaki I, Haidemenos G, et al. Melanoma/skin cancer screening in a Mediterranean country: results of the Euromelanoma Screening Day Campaign in Greece. Journal of the European Academy of Dermatology & Venereology 2007;21(1):56-62.

Stricklin 2011

Stricklin SM, Stoecker WV, Oliviero MC, Rabinovitz HS, Mahajan SK. Cloudy and starry milia-like cysts: how well do they distinguish seborrheic keratoses from malignant melanomas? Journal of the European Academy of Dermatology & Venereology 2011;25(10):1222-4.

Strumia 2003

Strumia R, Montanari A. Low positive predictive value of ABCD-E rule for dermatoscopy of small melanocytic naevi. Melanoma Research 2003;13:631-2.

Tasli 2012

Tasli L, Kacar N, Argenziano G. A scientometric analysis of dermoscopy literature over the past 25 years. Journal of the European Academy of Dermatology & Venereology 2012;26(9):1142-8.

Teban 2003

Teban L, Pehamberger H, Wolff K, Binder M, Kittler H. [Clinical value of a dermatoscopic classification of Clark nevi]. Journal der Deutschen Dermatologischen Gesellschaft 2003;1(4):292-6.

Terstappen 2007

Terstappen K, Larko O, Wennberg AM. Pigmented basal cell carcinoma--comparing the diagnostic methods of SIAscopy and dermoscopy. Acta Dermato-Venereologica 2007;87(3):238-42.

Terushkin 2010

Terushkin V, Braga JC, Dusza SW, Scope A, Busam K, Marghoob AA, et al. Agreement on the clinical diagnosis and management of cutaneous squamous neoplasms. Dermatologic Surgery 2010;36(10):1514-20.

Terushkin 2010a

Terushkin V, Warycha M, Levy M, Kopf AW, Cohen DE, Polsky D. Analysis of the benign to malignant ratio of lesions biopsied by a general dermatologist before and after the adoption of dermoscopy. Archives of Dermatology 2010;146(3):343-4.

Tromme 2012

Tromme I, Sacre L, Hammouch F, Legrand C, Marot L, Vereecken P, et al. Availability of digital dermoscopy in daily practice dramatically reduces the number of excised melanocytic lesions: results from an observational study. British Journal of Dermatology 2012;167(4):778-86.

Tschandl 2012

Tschandl P, Rosendahl C, Kittler H. Accuracy of the first step of the dermatoscopic 2-step algorithm for pigmented skin lesions. Dermatology Practical & Conceptual 2012;2(3):203a08.

Tschandl 2015

Tschandl P, Kittler H, Schmid K, Zalaudek I, Argenziano G. Teaching dermatoscopy of pigmented skin tumours to novices: comparison of analytic vs. heuristic approach. Journal of the European Academy of Dermatology & Venereology 2015;29(6):1198-204.

Tschandl 2015a

Tschandl P, Rosendahl C, Kittler H. Dermatoscopy of flat pigmented facial lesions. Journal of the European Academy of Dermatology & Venereology 2015;29(1):120-7.

Ulrich 2015

Ulrich M, von Braunmuehl T, Kurzen H, Dirschka T, Kellner C, Sattler E, et al. The sensitivity and specificity of optical coherence tomography for the assisted diagnosis of nonpigmented basal cell carcinoma: an observational study. British Journal of Dermatology 2015;173(2):428-35.

van der Leest 2011

van der Leest RJT, de Vries E, Bulliard JL, Paoli J, Peris K, Stratigos AJ, et al. The Euromelanoma skin cancer prevention campaign in Europe: characteristics and results of 2009 and 2010. Journal of the European Academy of Dermatology & Venereology 2011;25(12):1455-65.

van der Rhee 2010

van der Rhee JI, Bergman W, Kukutsch NA. The impact of dermoscopy on the management of pigmented lesions in everyday clinical practice of general dermatologists: a prospective study. British Journal of Dermatology 2010;162(3):563-7.

van der Rhee 2011

van der Rhee JI, Bergman W, Kukutsch NA. Impact of dermoscopy on the management of high-risk patients from melanoma families: a prospective study. Acta Dermato-Venereologica 2011;91(4):428-31.

Vasili 2010

Vasili E, Shkodrani E, Harja D, Labinoti L, Zoto A. Retrospective study of 70 patients with NMSC. Melanoma Research 2010;20:e63.

Verduzco-Martinez 2013

Verduzco-Martinez AP, Quinones-Venegas R, Guevara-Gutierrez E, Tlacuilo-Parra A. Correlation of dermoscopic findings with histopathologic variants of basal cell carcinoma. International Journal of Dermatology 2013;52(6):718-21.

Vestergaard 2008

Vestergaard ME, Macaskill P, Holt PE, Menzies SW. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. British Journal of Dermatology 2008;159(3):669-76.

Wang 2008

Wang SQ, Dusza SW, Scope A, Braun RP, Kopf AW, Marghoob AA. Differences in dermoscopic images from nonpolarized dermoscope and polarized dermoscope influence the diagnostic accuracy and confidence level: a pilot study. Dermatologic Surgery 2008;34(10):1389-95.

Warshaw 2009

Warshaw EM, Lederle FA, Grill JP, Gravely AA, Bangerter AK, Fortier LA, et al. Accuracy of teledermatology for pigmented neoplasms. Journal of the American Academy of Dermatology 2009;61(5):753-65.

Warshaw 2009a

Warshaw EM, Lederle FA, Grill JP, Gravely AA, Bangerter AK, Fortier LA, et al. Accuracy of teledermatology for nonpigmented neoplasms. Journal of the American Academy of Dermatology 2009;60(4):579-88.

Warshaw 2010

Warshaw EM, Gravely AA, Bohjanen KA, Chen K, Lee PK, Rabinovitz HS, et al. Interobserver accuracy of store and forward teledermatology for skin neoplasms. Journal of the American Academy of Dermatology 2010;62(3):513-6.

Weismann 2002

Weismann K, Lorentzen HF, Larsen FG. Diagnostic pearl: bright field globe magnifier diascopy for large pigmented skin lesions: a practical approach to epiluminescence microscopy. Journal of the American Academy of Dermatology 2002;47(2):304-6.

Wilkes 2010

Wilkes D. The use of dermoscopy in medical photography for the early detection of skin cancer. Journal of Visual Communication in Medicine 2010;33(4):169-73.

Winkelmann 2015

Winkelmann RR, Hauschild A, Tucker N, White R, Rigel DS. The Impact of Multispectral Digital Skin Lesion Analysis on German Dermatologist Decisions to Biopsy Atypical Pigmented Lesions with Clinical Characteristics of Melanoma. The Journal of Clinical & Aesthetic Dermatology 2015;8(10):27-9.

Winkelmann 2015a

Winkelmann RR, Yoo J, Tucker N, White R, Rigel DS. Impact of Guidance Provided by a Multispectral Digital Skin Lesion Analysis Device Following Dermoscopy on Decisions to Biopsy Atypical Melanocytic Lesions. The Journal of Clinical & Aesthetic Dermatology 2015;8(9):21-4.

Witkowski 2016

Witkowski AM, Ludzik J, DeCarvalho N, Ciardo S, Longo C, DiNardo A, et al. Non-invasive diagnosis of pink basal cell carcinoma: how much can we rely on dermoscopy and reflectance confocal microscopy? Skin Research & Technology 2016;22(2):230-7.

Yadav 1993

Yadav S, Vossaert KA, Kopf AW, Silverman M, Grin-Jorgensen C. Histopathologic correlates of structures seen on dermoscopy (epiluminescence microscopy). American Journal of Dermatopathology 1993;15(4):297-305.

Yamaura 2005

Yamaura M, Takata M, Miyazaki A, Saida T. Specific dermoscopy patterns and amplifications of the cyclin D1 gene to define histopathologically unrecognizable early lesions of acral melanoma in situ. Archives of Dermatology 2005;141(11):1413-8.

Yelamos 2016

Yelamos O, Nehal KS. Integrating clinical information, dermoscopy and reflectance confocal microscopy to improve the diagnostic accuracy and confidence of amelanotic and lightly pigmented melanomas. British Journal of Dermatology 2016;175(6):1147-8.

Yoo 2015

Yoo J, Tucker N, White R, Rigel D. The impact of probability of melanoma information provided by a multispectral digital skin lesion analysis device (MSDSLA) on resident dermatologists' decisions to biopsy clinical atypical lesions. Journal of the American Academy of Dermatology 2015;1):AB177.

Youl 2007

Youl PH, Raasch BA, Janda M, Aitken JF. The effect of an educational programme to improve the skills of general practitioners in diagnosing melanocytic/pigmented lesions. Clinical and Experimental Dermatology 2007;32(4):365-70.

Youl 2007a

Youl PH, Baade PD, Janda M, Del Mar CB, Whiteman DC, Aitken JF. Diagnosing skin cancer in primary care: how do mainstream general practitioners compare with primary care skin cancer clinic doctors? Medical Journal of Australia 2007;187(4):215-20.

Zaballos 2013

Zaballos P, Banuls J, Cabo H, Llambrich A, Salsench E, Puig S, et al. The usefulness of dermoscopy for the recognition of basal cell carcinoma--seborrhoeic keratosis compound tumours. Australasian Journal of Dermatology 2013;54(3):208-12.

Zalaudek 2010

Zalaudek I, Argenziano G, Marghoob AA, Pellacani G, Soyer HP. Dermoscopy and skin cancer. Dermatology research & Practice 2010;2010:867059.

Zell 2008

Zell D, Kim N, Olivero M, Elgart G, Rabinovitz H. Early diagnosis of multiple primary amelanotic/hypomelanotic melanoma using dermoscopy. Dermatologic Surgery 2008;34(9):1254-7.

Zortea 2014

Zortea M, Schopf TR, Thon K, Geilhufe M, Hindberg K, Kirchesch H, et al. Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists. Artificial Intelligence in Medicine 2014;60(1):13-26.

Zou 2001

Zou KH. Comparison of correlated receiver operating characteristic curves derived from repeated diagnostic test data. Academic Radiology 2001;8(3):225-33.

Studies awaiting classification

Ongoing studies

Other references

Additional references

ACIM 2014

Australian Cancer Database. Australian Cancer Incidence and Mortality (ACIM) books: Melanoma of the skin for Australia (ICD10 C43), 2011. www.aihw.gov.au/acim-books/ (accessed prior to 29 May 2018).

Ahmadi 2017

Ahmadi K, Prickaerts E, Smeets JGE, Joosten VHMJ, Kelleners-Smeets NWJ, Dinant GJ. Current approach of skin lesions suspected of malignancy in general practice in the Netherlands: a quantitative overview. Journal of the European Academy of Dermatology and Venereology 2017 Jul 27 [Epub ahead of print]. [DOI: 10.1111/jdv.14484; PubMed:

28750138]

Aldridge 2013

Aldridge RB, Naysmith L, Ooi ET, Murray CS, Rees JL. The importance of a full clinical examination: assessment of index lesions referred to a skin cancer clinic without a total body skin examination would miss one in three melanomas. Acta Dermato-Venereologica 2013;93(6):689-92. [PubMed: 23695107]

Altamura 2008

Altamura D, Avramidis M, Menzies SW. Assessment of the optimal interval for and sensitivity of short-term sequential digital dermoscopy monitoring for the diagnosis of melanoma. Archives of Dermatology 2008;144(4):502-6. [PubMed: 18427044]

Argenziano 2001

Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions--a valuable tool for early diagnosis of melanoma. Lancet Oncology 2001;2(7):443-9. [PubMed: 11905739]

Armstrong 2017

Armstrong BK, Cust AE. Sun exposure and skin cancer, and the puzzle of cutaneous melanoma: A perspective on Fears et al. Mathematical models of age and ultraviolet effects on the incidence of skin cancer among whites in the United States. American Journal of Epidemiology 1977; 105: 420-427. Cancer Epidemiology 2017;48:147-56. [PubMed: 28478931]

Arnold 2014

Arnold M, Holterhues C, Hollestein LM, Coebergh JW, Nijsten T, Pukkala E, et al. Trends in incidence and predictions of cutaneous melanoma across Europe up to 2015. Journal of the European Academy of Dermatology & Venereology 2014;28(9):1170-8. [PubMed: 23962170]

BAD 2013

British Association of Dermatology. Quality standards for Teledermatology using 'store and forward' images. www.bad.org.uk/shared/get-file.ashx?itemtype=document&id=794 (accessed prior to 29 May 2018).

Balch 2001

Balch CM, Soong SJ, Gershenwald JE, Thompson JF, Reintgen DS, Cascinelli N, et al. Prognostic factors analysis of 17,600 melanoma patients: validation of the American Joint Committee on Cancer melanoma staging system. Journal of Clinical Oncology 2001;19(16):3622-34. [PubMed: 11504744]

Balch 2009

Balch CM, Gershenwald JE, Soong SJ, Thompson JF, Atkins MB, Byrd DR, et al. Final Version of 2009 AJCC Melanoma Staging and Classification. Journal of Clinical Oncology 2009;27(36):6199-206. [PubMed: 19917835]

Belbasis 2016

Belbasis L, Stefanaki I, Stratigos AJ, Evangelou E. Non-genetic risk factors for cutaneous melanoma and keratinocyte skin cancers: An umbrella review of meta-analyses. Journal of Dermatological Science 2016;84(3):330-9. [PubMed: 27663092]

Boniol 2012

Boniol M, Autier P, Boyle P, Gandini S. Cutaneous melanoma attributable to sunbed use: systematic review and meta-analysis. BMJ 2012;345:e4757. [PubMed: 22833605]

Boring 1994

Boring CC, Squires TS, Tong T, Montgomery S. Cancer statistics, 1994. CA: a Cancer Journal for Clinicians 1994;44(1):7-26. [PubMed: 8281473]

Bossuyt 2015

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:h5527. [DOI: 10.1136/bmj.h5527; PubMed: 26511519]

Cancer Research UK 2017a

Cancer Research UK. Skin cancer statistics. www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer#heading-One (accessed prior to 19 July 2017).

Cancer Research UK 2017b

Cancer Research UK. Skin cancer incidence statistics. www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer/incidence (accessed prior to 19 July 2017).

Chao 2013

Chao D, London Cancer North and East. London Cancer, Guidelines for Cutaneous Malignant Melanoma Management August 2014. www.londoncancer.org/media/76373/london-cancer-melanoma-guidelines-2013-v1.0.pdf (accessed 25 February 2015).

Chappuis 2016

Chappuis P, Duru G, Marchal O, Girier P, Dalle S, Thomas L. Dermoscopy, a useful tool for general practitioners in melanoma screening: a nationwide survey. British Journal of Dermatology 2016;175(4):744-750. [PubMed: 26914613]

Cho 2014

Cho H, Mariotto AB, Schwartz LM, Luo J, Woloshin S. When do changes in cancer survival mean progress? The insight from population incidence and mortality. Journal of the National Cancer Institute. Monographs 2014;2014(49):187-97. [PubMed: 25417232]

Chu 2006

Chu H, Cole SR. Bivariate meta-analysis for sensitivity and specificity with sparse data: a generalized linear mixed model approach (comment). Journal of Clinical Epidemiology 2006;59(12):1331-2. [PubMed: 17098577]

Chuchu 2018a

Chuchu N, Dinnes J, Takwoingi Y, Matin RN, Bayliss SE, Davenport C, et al. The use of teledermatology for the diagnosis of skin cancer in adults. Cochrane Database of Systematic Reviews (in press).

Chuchu 2018b

Chuchu N, Takwoingi Y, Dinnes J, Matin RN, Bassett O, Moreau JF, et al. Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma. Cochrane Database of Systematic Reviews (in press).

Corbo 2012

Corbo MD, Wismer J. Agreement between dermatologists and primary care practitioners in the diagnosis of malignant melanoma: review of the literature. Journal of Cutaneous Medicine & Surgery 2012;16(5):306-10. [PubMed: 22971304]

Deeks 2005

Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. Journal of Clinical Epidemiology 2005;58(9):882-93. [PubMed: 16085191]

DePry 2011

DePry JL, Reed KB, Cook-Norris RH, Brewer JD. Iatrogenic immunosuppression and cutaneous malignancy [Review]. Clinics in Dermatology 2011;29(6):602-13. [PubMed: 22014982]

Dinnes 2018a

Dinnes J, Deeks JJ, Grainge MJ, Chuchu N, Ferrante di Ruffano L, Matin RN, et al. Visual inspection for the diagnosis of cutaneous melanoma in adults. Cochrane Database of Systematic Reviews (in press).

Dinnes 2018b

Dinnes J, Deeks JJ, Saleh D, Chuchu N, Bayliss SE, Patel L, et al. Reflectance confocal microscopy for the diagnosis of cutaneous melanoma in adults. Cochrane Database of Systematic Reviews (in press).

Dinnes 2018c

Dinnes J, Bamber J, Chuchu N, Bayliss SE, Takwoingi Y, Davenport C, et al. High frequency ultrasound for the diagnosis of skin cancer in adults. Cochrane Database of Systematic Reviews (in press).

Dinnes 2018d

Dinnes J, Deeks JJ, Chuchu N, Matin RN, Wong KY, Aldridge RB, et al. Visual examination and dermoscopy, alone or in combination, for the diagnosis of keratinocyte skin cancers in adults. Cochrane Database of Systematic Reviews (in press).

Efron 1983

Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American Statistical Association 1983;78(382):316-31. [DOI: 10.1080/01621459.1983.10477973]

Elstein 2002

Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. BMJ (Clinical Research Ed.) 2002;324(7339):729-32. [PubMed: 11909793]

Erdmann 2013

Erdmann F, Lortet-Tieulent J, Schuz J, Zeeb H, Greinert R, Breitbart EW, et al. International trends in the incidence of malignant melanoma 1953-2008--are recent generations at higher or lower risk? International Journal of Cancer 2013;132(2):385-400. [PubMed: 22532371]

EUCAN 2012

EUCAN, International Agency for Research on Cancer. Malignant melanoma of skin: Estimated incidence, mortality & prevalence for both sexes, 2012. eco.iarc.fr/eucan/Cancer.aspx?Cancer=20 (accessed 29 July 2015).

Ferlay 2015

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. International Journal of Cancer 2015;136(5):E359-86. [PubMed: 25220842]

Ferrante di Ruffano 2018a

Ferrante di Ruffano L, Dinnes J, Deeks JJ, Chuchu N, Bayliss SE, Davenport C, et al. Optical coherence tomography for the diagnosis of skin cancer in adults. Cochrane Database of Systematic Reviews (in press).

Ferrante di Ruffano 2018b

Ferrante di Ruffano L, Takwoingi Y, Dinnes J, Chuchu N, Bayliss SE, Davenport C, et al. Computer assisted diagnosis techniques (dermoscopy and spectroscopy-based) for the diagnosis of skin cancer in adults. Cochrane Database of Systematic Reviews (in press).

Friedman 1985

Friedman RJ, Rigel DS. The clinical features of malignant melanoma. Dermatologic Clinics 1985;3(2):271-83. [PubMed: 3830490]

Gandini 2005

Gandini S, Sera F, Cattaruzza MS, Pasquini P, Abeni D, Boyle P, et al. Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi [Review]. European Journal of Cancer 2005;41(1):28-44. [PubMed: 15617989]

Garbe 2016

Garbe C, Peris K, Hauschild A, Saiag P, Middleton M, Bastholt L, et al. Diagnosis and treatment of melanoma. European consensus-based interdisciplinary guideline - Update 2016. European Journal of Cancer 2016;63:201-17. [PubMed: 27367293]

Geller 2002

Geller AC, Miller DR, Annas GD, Demierre MF, Gilchrest BA, Koh HK. Melanoma incidence and mortality among US whites, 1969-1999. JAMA 2002;288(14):1719-20. [PubMed: 12365954]

Gershenwald 2017

Gershenwald JE, Scolyer RA, Hess KR, Sondak VK, Long GV, Ross, MI, et al. Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA: A Cancer Journal for Clinicians 2017;67(6):472-492. [PubMed: 29028110]

Grob 1998

Grob JJ, Bonerandi JJ. The 'ugly duckling' sign: identification of the common characteristics of nevi in an individual as a basis for melanoma screening. Archives of Dermatology 1998;134(1):103-4. [PubMed: 9449921]

Harrington 2017

Harrington E, Clyne B, Wesseling N, Sandhu H, Armstrong L, Bennett H, et al. Diagnosing malignant melanoma in ambulatory care: a systematic review of clinical prediction rules. BMJ Open 2017;7(3):e014096. [PubMed: 28264830]

HPA and MelNet NZ 2014

Health Promotion Agency and the Melanoma Network of New Zealand (MelNet). New Zealand Skin Cancer Primary Prevention and Early Detection Strategy 2014 to 2017. www.sunsmart.org.nz//sites/default/files/documents/NZ%20Skin%20Cancer%20PrimaryPrevention%20and%20EarlyDetection%20Strategy%202014%20to%202017%20FINAL%20VERS (accessed 29 May 2018).

Kasprzak 2015

Kasprzak JM, Xu YG. Diagnosis and management of lentigo maligna: a review. Drugs in Context 2015;4:212281. [PubMed: 26082796]

Kittler 2011

Kittler H, Rosendahl C, Cameron A, Tschandl P. Dermatoscopy. An algorithmic method based on pattern analysis. Austria: Facultas.WUV, 2011. [Other: ISBN-10: 3708907175]

Korn 2008

Korn EL, Liu PY, Lee SJ, Chapman JA, Niedzwiecki D, Suman VJ, et al. Meta-analysis of phase II cooperative group trials in metastatic stage IV melanoma to determine progression-free and overall survival benchmarks for future phase II trials. Journal of Clinical Oncology 2008;26(4):527-34. [PubMed: 18235113]

Kreusch 1991

Kreusch J, Rassner G. Standardized differentiation of melanocytic and non-melanocytic pigmented lesions by means of epiluminescent microscopy. Der Hautarzt; Zeitschrift fur Dermatologie, Venerologie, und Verwandte Gebiete 1991;42(2):77-83. [PubMed: 1828065]

Lachs 1992

Lachs MS, Nachamkin I, Edelstein PH, Goldman J, Feinstein AR, Schwartz JS. Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. Annals of Internal Medicine 1992;117(2):135-40. [PubMed: 1605428]

Leeflang 2013

Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. CMAJ 2013;185(11):E537-44. [PubMed: 23798453]

Leff 2008

Leff B, Finucane TE. Gizmo idolatry. JAMA 2008;299(15):1830-2. [PubMed: 18413879]

Lehmann 2011

Lehmann AR, McGibbon D, Stefanini M. Xeroderma pigmentosum. Orphanet Journal Of Rare Diseases 2011;6:70. [PubMed: 22044607]

Linos 2009

Linos E, Swetter SM, Cockburn MG, Colditz GA, Clarke CA. Increasing burden of melanoma in the United States. Journal of Investigative Dermatology 2009;129(7):1666-74. [PubMed: 19131946]

Loescher 2011

Loescher LJ, Harris JM Jr, Curiel-Lewandrowski C. A systematic review of advanced practice nurses' skin cancer assessment barriers,skin lesion recognition skills,and skin cancer training activities. Journal of the American Academy of Nurse Practitioners 2011;23(12):667-73. [PubMed: 22145657]

MacKie 1985

MacKie RM, English J, Aitchison TC, Fitzsimons CP, Wilson P. The number and distribution of benign pigmented moles (melanocytic naevi) in a healthy British population. British Journal of Dermatology 1985;113(2):167-74. [PubMed: 4027184]

MacKie 1990

MacKie RM. Clinical recognition of early invasive malignant melanoma. BMJ 1990;301(6759):1005-6. [PubMed: 2249043]

Maley 2014

Maley A, Rhodes AR. Cutaneous melanoma: preoperative tumor diameter in a general dermatology outpatient setting. Dermatologic Surgery 2014;40(4):446-54. [PubMed: 24479783]

Marsden 2010

Marsden JR, Newton-Bishop JA, Burrows L, Cook M, Corrie PG, Cox NH, et al. BAD Guidelines: Revised UK guidelines for the management of cutaneous melanoma 2010. British Journal of Dermatology 2010;163(2):238-56. [PubMed: 20608932]

Mistry 2011

Mistry M, Parkin DM, Ahmad AS, Sasieni P. Cancer incidence in the United Kingdom: projections to the year 2030. British Journal of Cancer 2011;105(11):1795-803. [PubMed: 22033277]

Moons 1997

Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. Epidemiology 1997;8(1):12-7. [PubMed: 9116087]

Ndegwa 2010

Ndegwa S, Prichett-Pejic W, McGill S, Murphy G, Severn M. Teledermatology services: rapid review of diagnostic, clinical management, and economic outcomes. www.cadth.ca/media/pdf/H0502_Teledermatology_Report_e.pdf (accessed prior to 29 May 2018).

NICE 2015a

National Institute for Health and Care Excellence. Melanoma: assessment and management. www.nice.org.uk/guidance/ng14 (accessed prior to 19 July 2017).

NICE 2015b

National Institute for Health and Clinical Excellence. Suspected cancer: recognition and referral. www.nice.org.uk/guidance/ng12 (accessed prior to 28 March 2018).

Norman 1989

Norman GR, Rosenthal D, Brooks LR, Allen SW, Muzzin LJ. The development of expertise in dermatology. Archives of Dermatology 1989;125(8):1063-8. [PubMed: 2757402]

Norman 2009

Norman G, Barraclough K, Dolovich L, Price D. Iterative diagnosis. BMJ 2009;339:b3490. [PubMed: 19773326]

Pasquali 2018

Pasquali S, Hadjinicolaou AV, Chiarion Sileni V, Rossi CR, Mocellin S. Systemic treatments for metastatic cutaneous melanoma. Cochrane Database of Systematic Reviews 2018, Issue 2. Art. No.: CD011123 DOI: 10.1002/14651858.CD011123.pub2.

Pehamberger 1993

Pehamberger H, Binder M, Steiner A, Wolff K. In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma. Journal of Investigative Dermatology 1993;100(3):356s-62s. [PubMed: 8440924]

Reitsma 2005

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinical Epidemiology 2005;58(10):982-90. [PubMed: 16168343]

Reyes-Ortiz 2006

Reyes-Ortiz CA, Goodwin JS, Freeman JL, Kuo YF. Socioeconomic status and survival in older patients with melanoma. Journal of the American Geriatrics Society 2006;54(11):1758-64. [PubMed: 17087705]

Rozeman 2017

Rozeman EA, Dekker TJA, Haanen JBAG, Blank CU. Advanced Melanoma: Current Treatment Options, Biomarkers, and Future Perspectives. American Journal of Clinical Dermatology 2017 Nov 21 [Epub ahead of print]. [DOI: 10.1007/s40257-017-0325-6; PubMed: 29164492]

Rutjes 2005

Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. Clinical Chemistry 2005;51(8):1335-41. [PubMed: 15961549]

Rutjes 2006

Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006;174(4):469-76. [PubMed: 16477057]

Rutter 2001

Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Statistics in Medicine 2001;20(19):2865-84. [PubMed: 11568945]

SAS 2012

SAS 2012 [Computer program]. Version 9.3. Cary, NC, USA: SAS Institute Inc., 2012.

Scope 2006

Scope A, Burroni M, Agero AL, Benvenuto-Andrade C, DuszaSW, Rubegni P, et al. Predominant dermoscopic patterns observed among nevi. Journal of Cutaneous Medicine and Surgery 2006;10(4):170-4. [PubMed: 17234115]

Siegel 2015

Siegel R, Miller K, Jemal A. Cancer statistics, 2015. CA: a Cancer Journal for Clinicians 2015;65(1):5-29. [PubMed: 25559415]

SIGN 2017

Scottish Intercollegiate Guidelines Network. Cutaneous Melanoma. www.sign.ac.uk/sign-146-melanoma.html (accessed prior to 19 July 2017).

Sladden 2009

Sladden MJ, Balch C, Barzilai DA, Berg D, Freiman A, Handiside T, et al. Surgical excision margins for primary cutaneous melanoma. Cochrane Database of Systematic Reviews 2009, Issue 10. Art. No.: CD004835 DOI: 10.1002/14651858.CD004835.pub2.

Slater 2014

Slater D, Walsh M. Standards and datasets for reporting cancers: Dataset for the histological reporting of primary cutaneous malignant melanoma and regional lymph nodes, May 2014. www.rcpath.org/Resources/RCPath/Migrated%20Resources/Documents/G/G125_DatasetMaligMelanoma_May14.pdf (accessed 29 July 2015).

Sober 1979

Sober AJ, Fitzpatrick TB, Mihm MC, Wise TG, Pearson BJ, Clark WH, et al. Early recognition of cutaneous melanoma. JAMA 1979;242(25):2795-9. [PubMed: 501893]

Swerdlow 1995

Swerdlow AJ, English JS, Qiao Z. The risk of melanoma in patients with congenital nevi: a cohort study. Journal of the American Academy of Dermatology 1995;32(4):595-9. [PubMed: 7896948]

Takwoingi 2010

Takwoingi Y, Deeks J. MetaDAS: a SAS macro for meta-analysis of diagnostic accuracy studies. User Guide Version 1.3. 2010. www.methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/MetaDAS%20Readme%20v1.3%20May%202012.pdf (accessed prior to 17 July 2017).

Takwoingi 2013

Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. Annals of Internal Medicine 2013;158(7):544-54. [PubMed: 23546566]

Takwoingi 2015

Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. Statistical Methods in Medical Research 2015;24:1-19. [DOI: 10.1177/0962280215592269]

Tucker 1985

Tucker MA, Boice JD Jr, Hoffman DA. Second cancer following cutaneous melanoma and cancers of the brain, thyroid, connective tissue, bone, and eye in Connecticut, 1935-82. National Cancer Institute Monographs 1985;68:161-89. [PubMed: 4088297]

Usher-Smith 2016

Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. BMJ (Clinical Research Ed.) 2016;353:i3139. [DOI: 10.1136/bmj.i3139; PubMed: 27334281]

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine 2011;155(8):529-36. [PubMed: 22007046]

Other published versions of this review

Dinnes 2015a

Dinnes J, Matin RN, Moreau JF, Patel L, Chan SA, Wong KY, et al. Tests to assist in the diagnosis of cutaneous melanoma in adults: a generic protocol. Cochrane Database of Systematic Reviews 2015, Issue 10. Art. No.: CD011902 DOI: 10.1002/14651858.CD011902.

Dinnes 2015b

Dinnes J, Wong KY, Gulati A, Chuchu N, Leonardi-Bee J, Bayliss SE, et al. Tests to assist in the diagnosis of keratinocyte skin cancers in adults: a generic protocol. Cochrane Database of Systematic Reviews 2015, Issue 10. Art. No.: CD011901 DOI: 10.1002/14651858.CD011901.

Classification pending references

Data and analyses

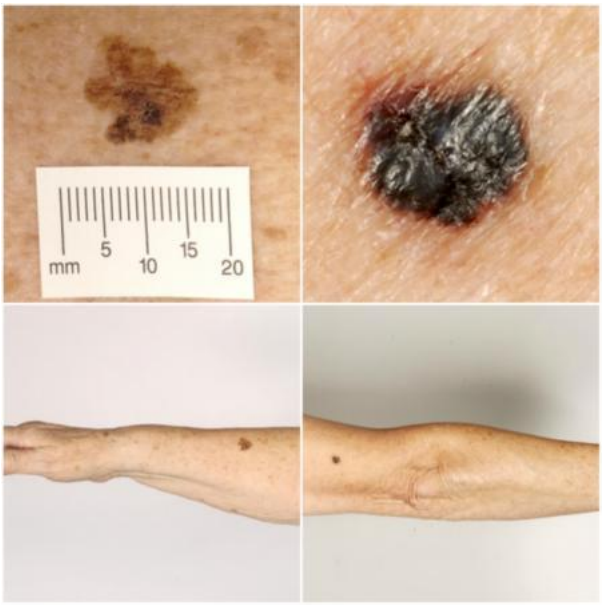
Data tables by test

Test	Studies	Participants
1 Visual inspection - in-person (MM)	2	147
2 VI+Dermoscopy - in-person (MM)	6	789
3 Visual inspection - image-based (MM)	4	454
4 Dermoscopy alone - image-based (MM)	13	5618
5 Visual inspection - in-person (MM+MiS)	13	6740
6 VI+Dermoscopy - in-person (MM+MiS)	26	23169
7 Visual inspection - image-based (MM+MiS)	11	1740
8 Dermoscopy alone - image-based (MM+MiS)	60	13475
9 Visual inspection - in-person (Any)	2	3457
10 VI+Dermoscopy - in-person (Any)	4	3880
11 Visual inspection - image-based (Any)	3	547
12 Dermoscopy alone - image-based (Any)	5	815
13 MM2- Dermoscopy - no algorithm - threshold NR - in-person	8	4707
14 MM2- Dermoscopy - pattern - at >=1 char present - in-person	1	220
15 MM2- Dermoscopy - pattern - at >=3 chars present - in-person	1	68
16 MM2- Dermoscopy - pattern - threshold NR - in-person	6	4307
17 MM2- Dermoscopy - ABCD at NR (likely >5.45) - in-person	1	235
18 MM2- Dermoscopy - ABCD at >5.45 - in-person	4	1203
19 MM2- Dermoscopy - ABCD at >4.75 - in-person	1	309
20 MM2- Dermoscopy - ABCD at 60% specificity - in-person	1	356
21 MM2- Dermoscopy - ABCD at 80% specificity - in-person	1	356
22 MM2- Dermoscopy - ABCD at 70% specificity - in-person	1	356
23 MM2- Dermoscopy - ABCD at 75% specificity - in-person	1	356
24 MM2- Dermoscopy - ABCD at 85% specificity - in-person	1	356
25 MM2- Dermoscopy - ABCD at 90% specificity - in-person	1	356
26 MM2- Dermoscopy - ABCDE at >1.3 - in-person	1	356
27 MM2- Dermoscopy - ABCDE at >2.65 - in-person	1	356
28 MM2- Dermoscopy - ABCDE at >3.05 - in-person	1	356
29 MM2- Dermoscopy - ABCDE at >3.6 - in-person	1	356
30 MM2- Dermoscopy - ABCDE at >4.25 - in-person	1	356
31 MM2- Dermoscopy - ABCDE at >4.9 - in-person	1	356
32 MM2- Dermoscopy - 7FFM at >=2 - in-person	1	401
33 MM2- Dermoscopy - 7point at >=2 - in-person	1	638
34 MM2- Dermoscopy - 7point at >=3 - in-person	2	11137
35 MM2- Dermoscopy - Menzies at 2 neg and >=1 pos - in-person	1	206
36 MM2- Dermoscopy - no algorithm - any threshold - image-based	24	4498
37 MM2- Dermoscopy - no algorithm - correct dx - image-based	18	4118
38 MM2- Dermoscopy - no algorithm - excise decision - image-based	10	831
39 MM2- Dermoscopy - pattern - any threshold - image-based	20	4621
40 MM2- Dermoscopy - pattern - correct dx - image-based	19	4095
41 MM2- Dermoscopy - pattern - excise decision - image-based	3	933
42 MM2- Dermoscopy - ABCD at >4.75 - image-based	10	4242
43 MM2- Dermoscopy - ABCD at >5.45 - image-based	7	2471
44 MM2- Dermoscopy - rev ABCD at >=4 - image-based	1	269
45 MM2- Dermoscopy - ABCDE at >=4 - image-based	1	269
46 MM2- Dermoscopy - 7point at NR - image-based	4	1936
47 MM2- Dermoscopy - 7point at >=3 - image-based	11	3408
48 MM2- Dermoscopy - 7point at >=5 - image-based	1	322
49 MM2- Dermoscopy - rev 7point at NR (likely >=1) - image-based	1	1678
50 MM2- Dermoscopy – rev 7point at >=1 - image-based	1	300
51 MM2- Dermoscopy - rev 7point for FU - major change - image-based	1	70
52 MM2- Dermoscopy - 7FFM at >=2 - image-based	4	2200
53 MM2- Dermoscopy - Menzies at 2neg and >=1 pos - image-based	4	1856
54 MM2- Dermoscopy - Menzies at NR - image-based	2	60

55 MM2- Dermoscopy - 3point at >=2 - image-based	7	1505
56 MM2- Dermoscopy - 4point (scored 3-point) at >2 - image-based	1	75
57 MM2- Dermoscopy - Hofman algorithm at NR - image-based	1	254
58 MM2- Dermoscopy CASH at >=6 - image-based	1	477
59 MM2- Dermoscopy CASH at >=8 - image-based	2	190
60 MM2- Dermoscopy Chaos/Clues at =2 - image-based	2	940
61 MM2- Dermoscopy - Acral 3step - image-based	1	107
62 VI+Dermoscopy (in-person) - observer experience NR (MM+MiS)	10	8390
63 VI+Dermoscopy (in-person) - high experience (MM+MiS)	14	14213
65 VI+Dermoscopy (in-person) - trained observer (MM+MiS)	2	566
66 Dermoscopy (image-based) - observer experience NR (MM+MiS)	11	2777
67 Dermoscopy (image-based) - high experience (MM+MiS)	34	8933
68 Dermoscopy (image-based) - moderate experience (MM+MiS)	5	678
69 Dermoscopy (image-based) - low experience (MM+MiS)	6	448
70 Dermoscopy (image-based) - mixed experience (MM+MiS)	5	473
71 Dermoscopy (image-based) - trained observer (MM+MiS)	11	1087
72 VI+Dermoscopy (in-person) - Consultant expert (MM+MiS)	11	2767
73 VI+Dermoscopy (in-person) - Consultant (MM+MiS)	10	8390
74 VI+Dermoscopy (in-person) - Resident/registrar (MM+MiS)	2	11137
75 VI+Dermoscopy (in-person) - Mixed (secondary care based) (MM+MiS)	1	309
76 VI+Dermoscopy (in-person) - GP (MM+MiS)	2	566
77 Dermoscopy (image-based) - Consultant expert (MM+MiS)	33	8664
78 Dermoscopy (image-based) - Consultant (MM+MiS)	24	3986
79 Dermoscopy (image-based) - Resident (MM+MiS)	5	927
80 Dermoscopy (image-based) - Mixed (secondary care based) (MM+MiS)	4	399
81 Dermoscopy (image-based) - Mixed (other) (MM+MiS)	4	867
82 Dermoscopy (image-based) - GP/primary care (MM+MiS)	3	288
83 Dermoscopy (image-based) - Physician assistant (MM+MiS)	1	65
84 Dermoscopy - before training (MM+MiS)	4	245
85 Dermoscopy - after training (MM+MiS)	4	245
86 Dermoscopy - before training (MM)	2	150
87 Dermoscopy - after training (MM)	2	150
88 MM1- Dermoscopy - no algorithm - threshold NR - in-person	3	190
89 MM1- Dermoscopy - pattern analysis - threshold NR - in-person	1	45
90 MM1- Dermoscopy - ABCD at >4.2 - in-person	1	495
91 MM1- Dermoscopy - ABCD at >5.45 - in-person	2	832
92 MM1- Dermoscopy - Kenet (modified) at melanoma possible - in-person	1	54
93 MM1- Dermoscopy - Kenet (modified) at melanoma likely - in-person	1	54
94 MM1- Dermoscopy - no algorithm - threshold NR - image-based	6	683
95 MM1- Dermoscopy - no algorithm - decision to excise - image-based (paired data only)	1	99
96 MM1- Dermoscopy - pattern analysis - threshold NR - image-based	1	119
97 MM1- Dermoscopy - ABCD at >4.75 - image-based	2	330
98 MM1- Dermoscopy - ABCD at >5.45 - image-based	1	258
99 MM1- Dermoscopy - 7point at NR - image-based	1	332
100 MM1- Dermoscopy - Menzies at 2neg and >=1 pos - image-based	4	4184
101 MM1- Dermoscopy - 3point at >NR - image-based	1	332
102 MM1- Dermoscopy - Kenet at melanoma likely - image-based	1	258
103 MM1- Dermoscopy - Kenet at melanoma possible - image-based	1	258
104 MM1- Dermoscopy CASH at >=8 - image-based	1	332
105 MM1- Dermoscopy - Kreusch algorithm - image-based	1	265
106 MM1- Dermoscopy - Menzies for amelanotic at 1 - image-based	1	332
107 MM1- Dermoscopy - Menzies for amelanotic at 0 - image-based	1	332
108 MM3- Dermoscopy – no algorithm at NR - in-person	1	231
109 MM3- Dermoscopy - pattern analysis - threshold NR - in-person	1	3372
110 MM3- Dermoscopy – ABCD at >5.45 - in-person	1	200
111 MM3- Dermoscopy – 3point at >=2 - in-person	1	77
112 MM3- Dermoscopy – no algorithm at NR - image-based	2	83
113 MM3- Dermoscopy - pattern analysis - threshold NR - image-based	1	119
114 MM3- Dermoscopy – 3point at >=2 - image-based	1	150
115 MM2 - VI - in-person (w image-based dermoscopy)	2	886

Figures

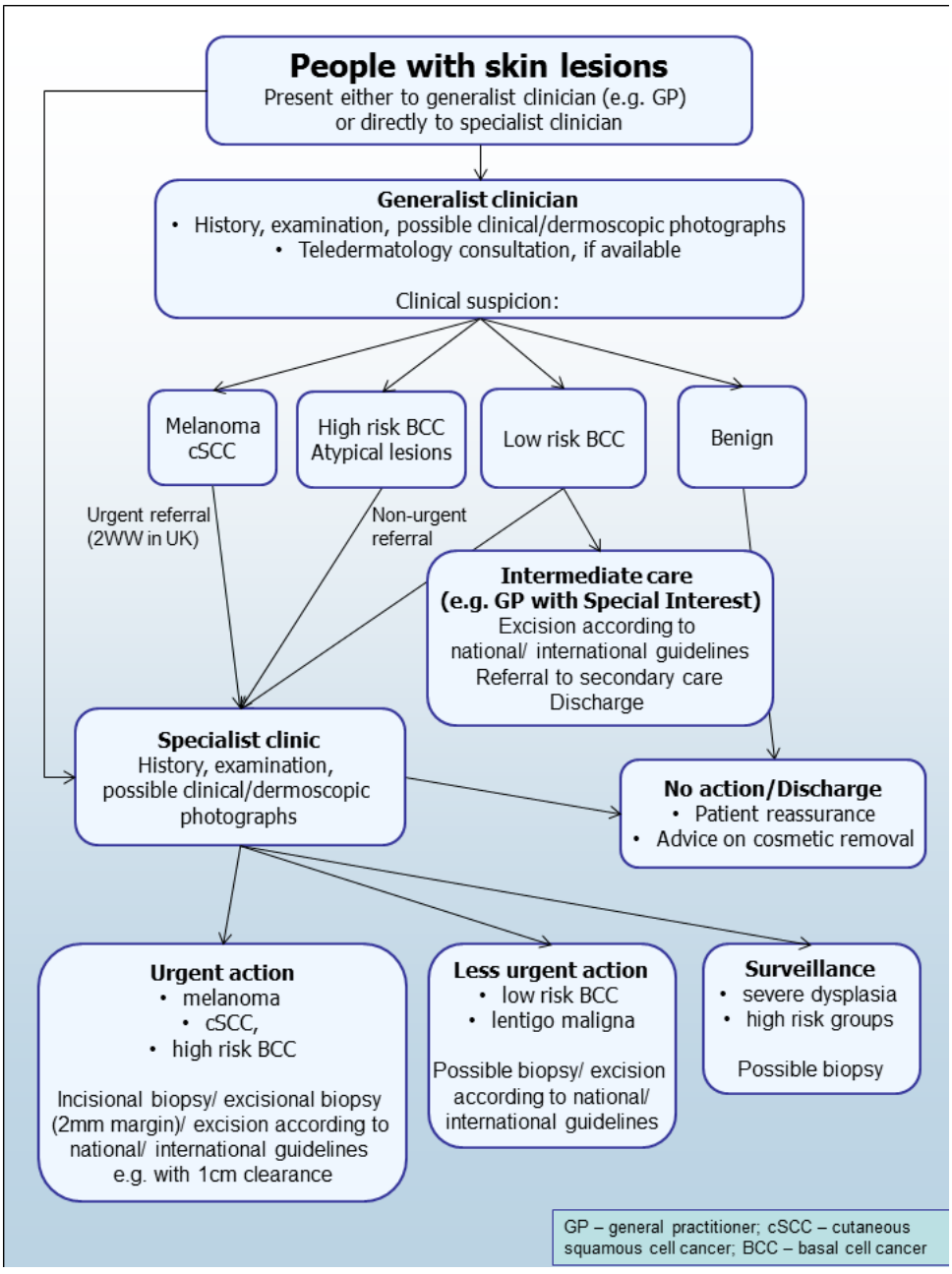
Figure 1



Caption

Sample photographs of superficial spreading melanoma (left) and nodular melanoma (right)

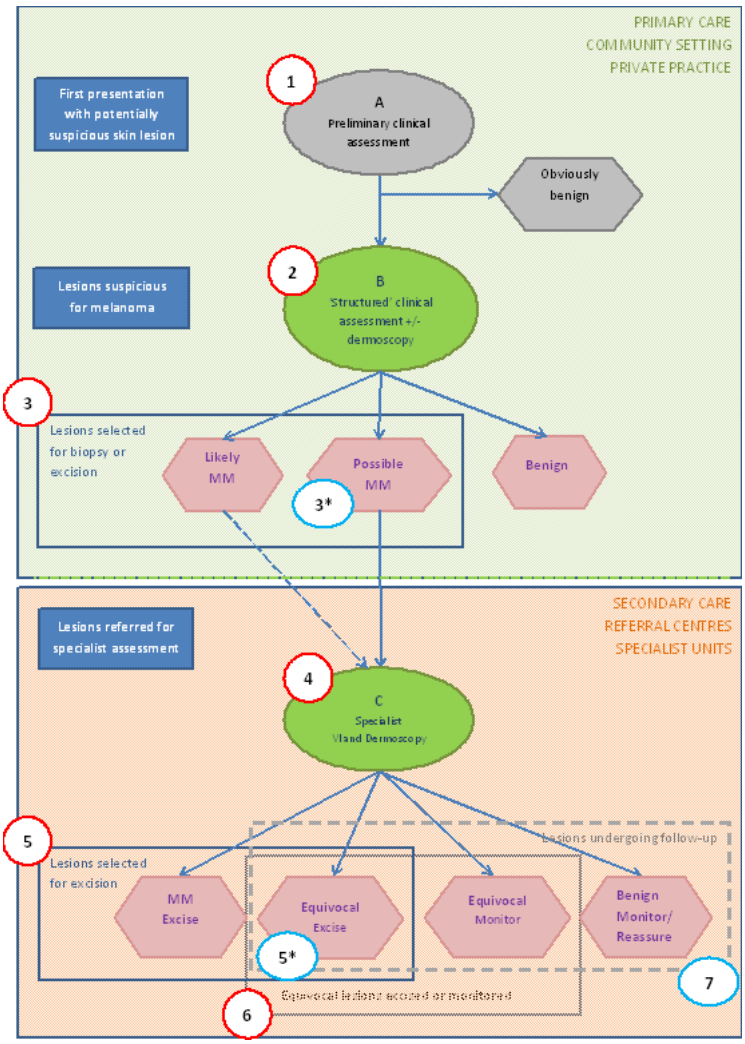
Figure 2



Caption

Current clinical pathway for people with skin lesions

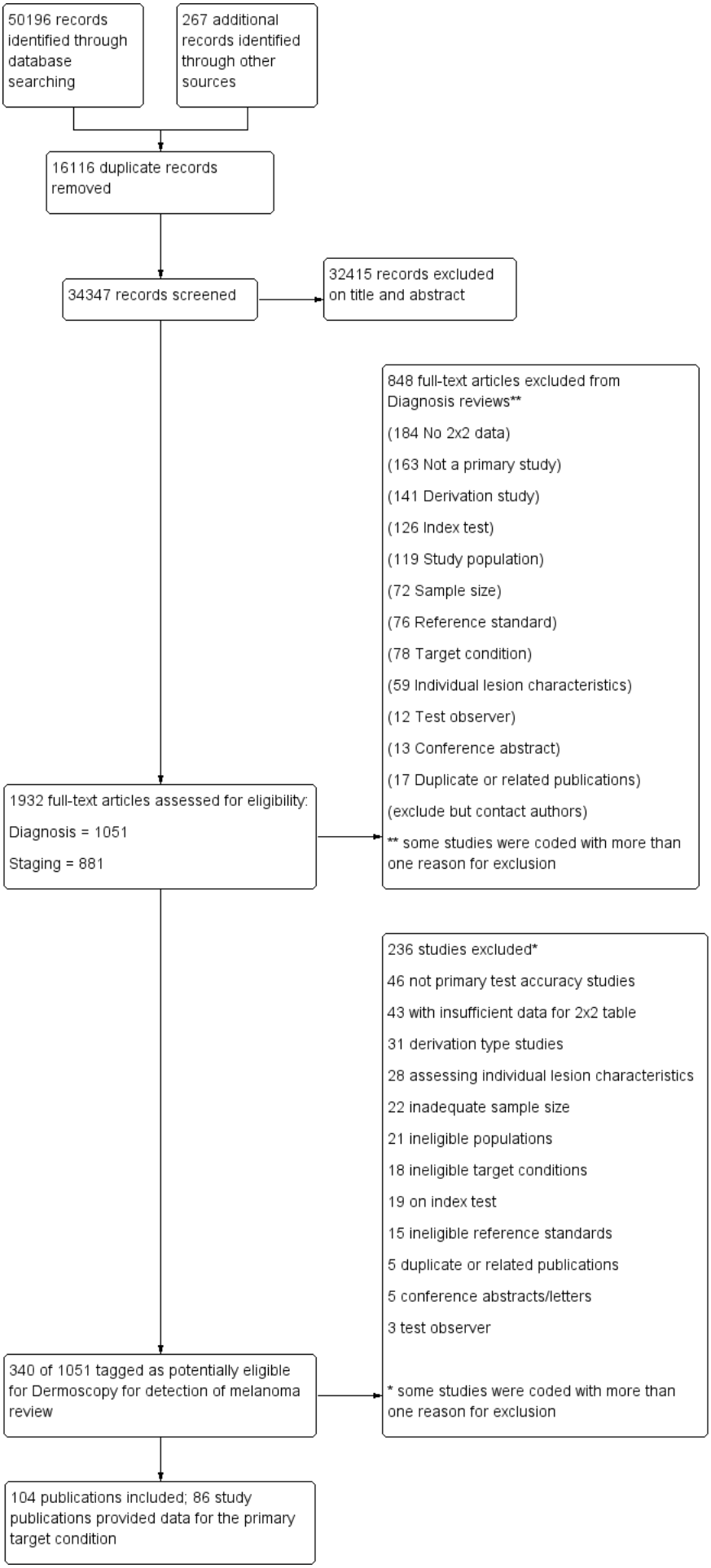
Figure 3



Caption

Clinical pathway

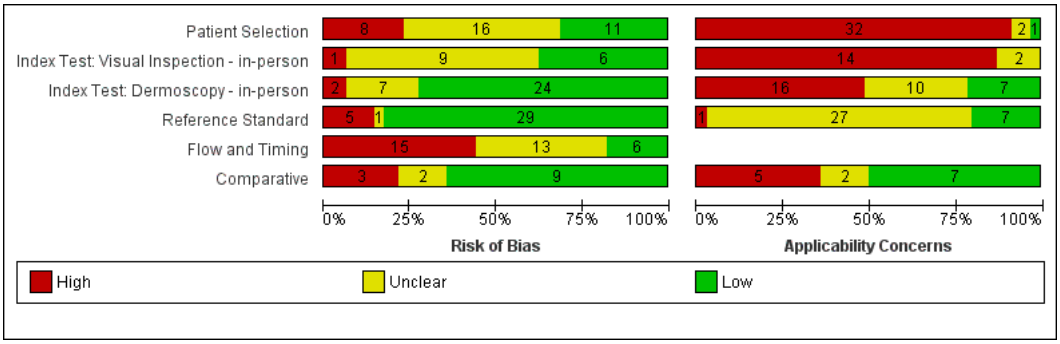
Figure 4



Caption

PRISMA flow diagram.

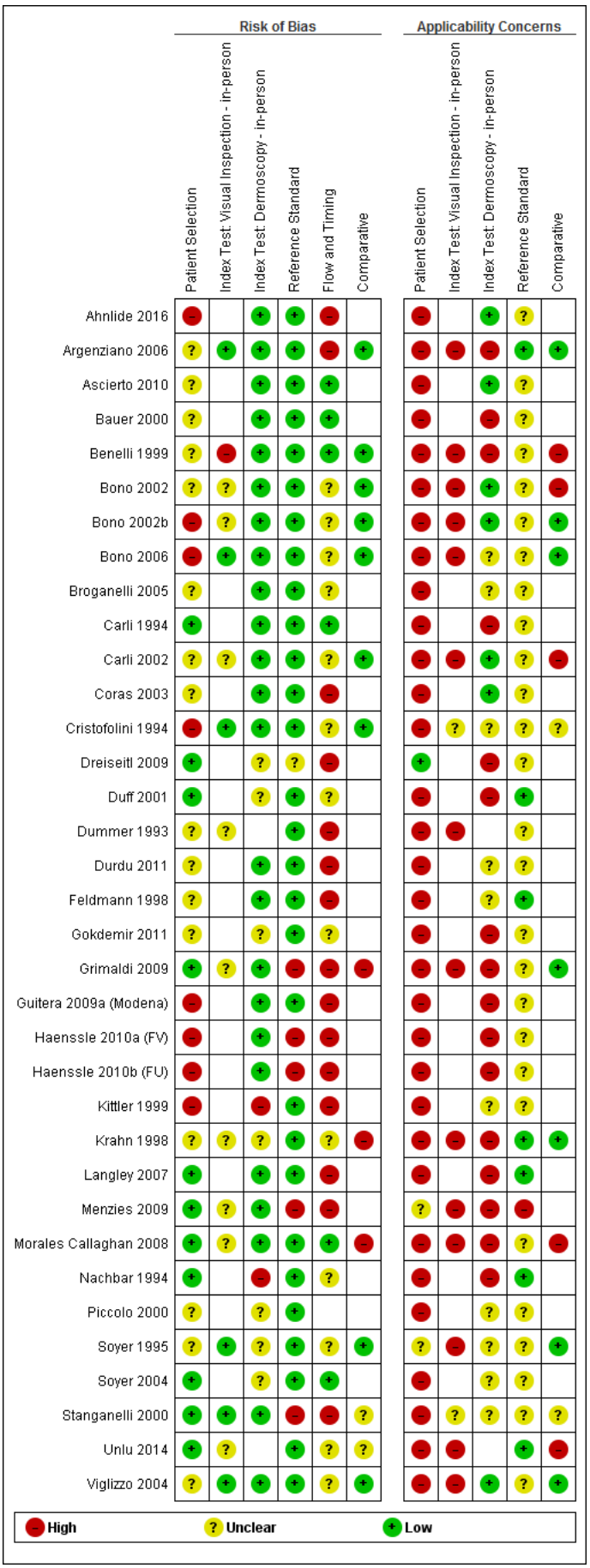
Figure 5



Caption

Risk of bias and applicability concerns graph for in-person evaluations: review authors' judgements about each domain presented as percentages across included studies

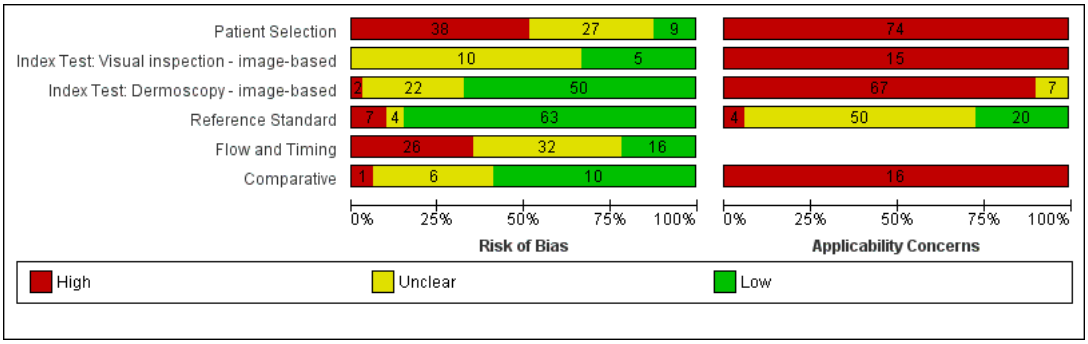
Figure 6



Caption

Risk of bias and applicability concerns for in-person evaluations summary: review authors' judgements about each domain for each included study

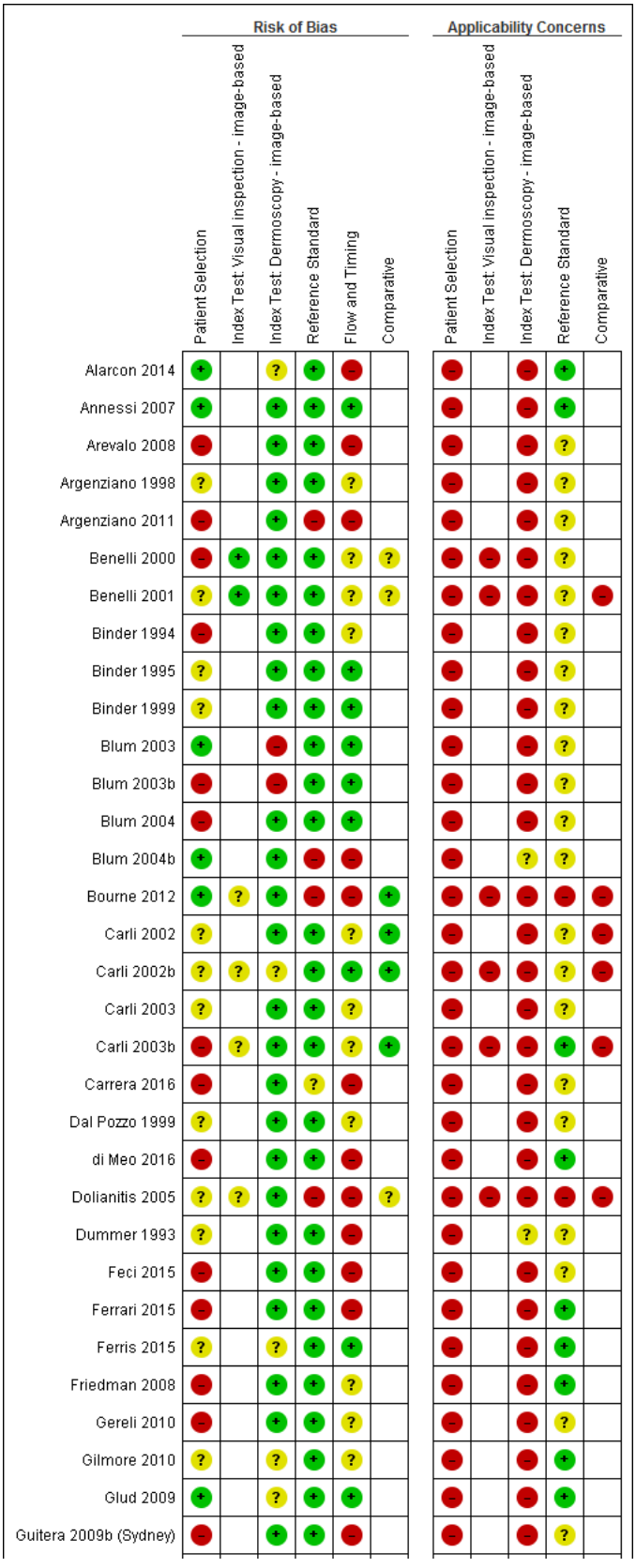
Figure 7

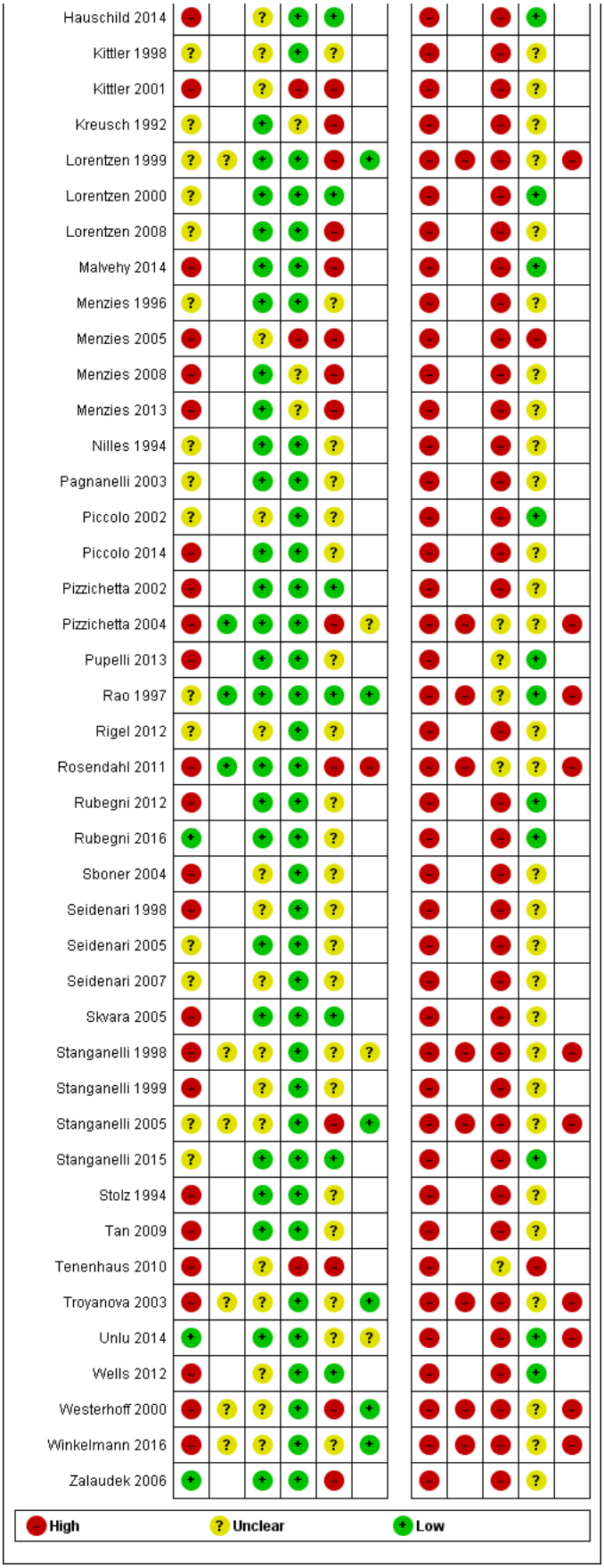


Caption

Risk of bias and applicability concerns graph for image-based evaluations: review authors' judgements about each domain presented as percentages across included studies

Figure 8



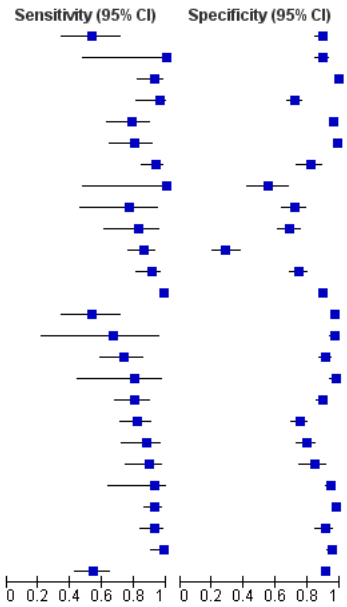


Caption

Risk of bias and applicability concerns for image-based evaluations summary: review authors' judgements about each domain for each included study

Figure 9 (Analysis 1)

Study	TP	FP	FN	TN	Pathway	Sensitivity (95% CI)	Specificity (95% CI)
Menzies 2009	17	33	15	266	2-c	0.53 [0.35, 0.71]	0.89 [0.85, 0.92]
Grimaldi 2009	5	24	0	206	2-c	1.00 [0.48, 1.00]	0.90 [0.85, 0.93]
Stanganelli 2000	51	9	4	3308	4-c	0.93 [0.82, 0.98]	1.00 [0.99, 1.00]
Dreiseitl 2009	26	121	1	310	4-c	0.96 [0.81, 1.00]	0.72 [0.67, 0.76]
Bauer 2000	33	10	9	263	4-u	0.79 [0.63, 0.90]	0.96 [0.93, 0.98]
Haenssle 2010a (FV)	32	146	8	8263	4-u	0.80 [0.64, 0.91]	0.98 [0.98, 0.99]
Soyer 1995	61	17	4	77	5*-c	0.94 [0.85, 0.98]	0.82 [0.73, 0.89]
Carli 1994	5	28	0	35	5*-c	1.00 [0.48, 1.00]	0.56 [0.42, 0.68]
Bono 2002b	10	42	3	106	5-c	0.77 [0.46, 0.95]	0.72 [0.64, 0.79]
Bono 2006	19	57	4	126	5-c	0.83 [0.61, 0.95]	0.69 [0.62, 0.75]
Guitera 2009a (Modena)	68	83	11	33	5-c	0.86 [0.76, 0.93]	0.28 [0.20, 0.38]
Bono 2002	60	63	6	184	5-c	0.91 [0.81, 0.97]	0.74 [0.69, 0.80]
Duff 2001	577	193	9	1593	5-c	0.98 [0.97, 0.99]	0.89 [0.88, 0.91]
Feldmann 1998	16	14	14	456	5-u	0.53 [0.34, 0.72]	0.97 [0.95, 0.98]
Morales Callaghan 2008	4	6	2	188	5-u	0.67 [0.22, 0.96]	0.97 [0.93, 0.99]
Ahnlide 2016	34	23	12	240	5-u	0.74 [0.59, 0.86]	0.91 [0.87, 0.94]
Durdu 2011	8	5	2	185	5-u	0.80 [0.44, 0.97]	0.97 [0.94, 0.99]
Benelli 1999	48	37	12	304	5-u	0.80 [0.68, 0.89]	0.89 [0.85, 0.92]
Kittler 1999	60	71	13	212	5-u	0.82 [0.71, 0.90]	0.75 [0.69, 0.80]
Cristofolini 1994	29	39	4	148	5-u	0.88 [0.72, 0.97]	0.79 [0.73, 0.85]
Langley 2007	33	14	4	74	5-u	0.89 [0.75, 0.97]	0.84 [0.75, 0.91]
Gokdemir 2011	12	25	1	410	5-u	0.92 [0.64, 1.00]	0.94 [0.92, 0.96]
Broganelli 2005	100	10	8	520	5-u	0.93 [0.86, 0.97]	0.98 [0.97, 0.99]
Nachbar 1994	64	11	5	114	5-u	0.93 [0.84, 0.98]	0.91 [0.85, 0.96]
Carli 2002	53	9	1	193	5-u	0.98 [0.90, 1.00]	0.96 [0.92, 0.98]
Haenssle 2010b (FU)	47	228	40	2373	7-u	0.54 [0.43, 0.65]	0.91 [0.90, 0.92]

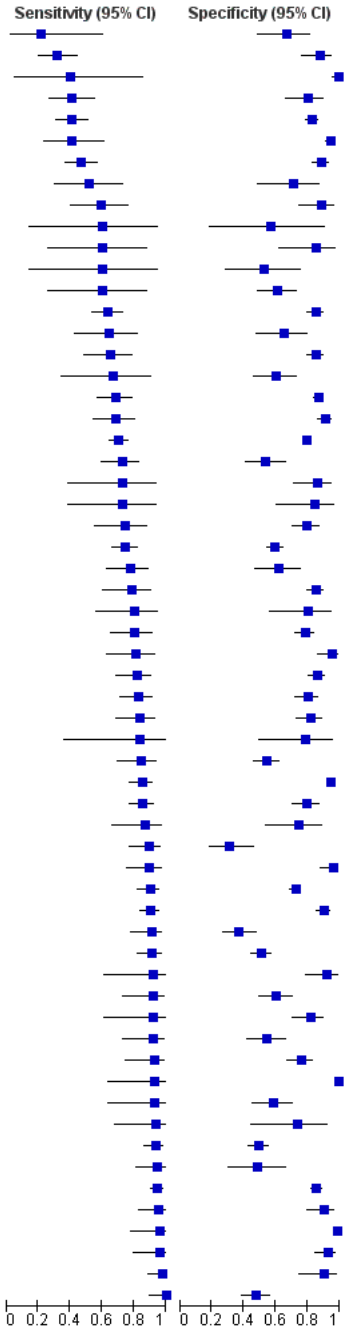


Caption

In-person evaluations of the accuracy of dermoscopy added to visual inspection grouped by pathway categorisation for detecting invasive melanoma or melanocytic intraepidermal variants

Figure 10 (Analysis 2)

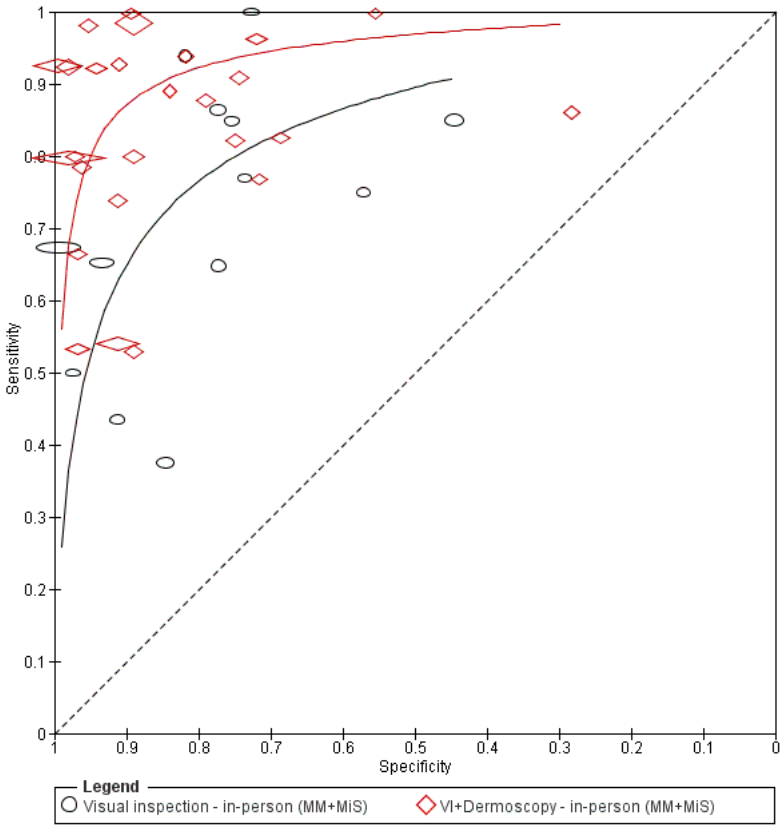
Study	TP	FP	FN	TN	Pathway	Sensitivity (95% CI)	Specificity (95% CI)
Bourne 2012	2	12	7	24	3-c	0.22 [0.03, 0.60]	0.67 [0.49, 0.81]
Skvara 2005	20	8	43	55	7-u	0.32 [0.21, 0.45]	0.87 [0.77, 0.94]
Pizzichetta 2002	2	1	3	123	5-c	0.40 [0.05, 0.85]	0.99 [0.96, 1.00]
Friedman 2008	20	10	29	40	5-u	0.41 [0.27, 0.56]	0.80 [0.66, 0.90]
Menzies 2008	43	67	62	325	5-u	0.41 [0.31, 0.51]	0.83 [0.79, 0.87]
Rosendahl 2011	12	25	17	409	3-u	0.41 [0.24, 0.61]	0.94 [0.92, 0.96]
Argenziano 2011	47	23	53	177	5-c	0.47 [0.37, 0.57]	0.89 [0.83, 0.93]
Wells 2012	12	7	11	17	5-u	0.52 [0.31, 0.73]	0.71 [0.49, 0.87]
di Meo 2016	19	5	13	38	5-u	0.59 [0.41, 0.76]	0.88 [0.75, 0.96]
Winkelmann 2016	3	3	2	4	5-u	0.60 [0.15, 0.95]	0.57 [0.18, 0.90]
Stanganelli 1998	6	3	4	17	5-u	0.60 [0.26, 0.88]	0.85 [0.62, 0.97]
Rigel 2012	3	9	2	10	5-u	0.60 [0.15, 0.95]	0.53 [0.29, 0.76]
Kittler 2001	6	27	4	43	7-u	0.60 [0.26, 0.88]	0.61 [0.49, 0.73]
Feci 2015	65	33	37	186	5-u	0.64 [0.54, 0.73]	0.85 [0.79, 0.89]
Ferris 2015	16	14	9	26	5-u	0.64 [0.43, 0.82]	0.65 [0.48, 0.79]
Seidenari 2007	28	30	15	170	5-u	0.65 [0.49, 0.79]	0.85 [0.79, 0.90]
Stanganelli 2015	8	23	4	35	7-u	0.67 [0.35, 0.90]	0.60 [0.47, 0.73]
Benelli 2000	52	70	24	454	5-u	0.68 [0.57, 0.79]	0.87 [0.83, 0.89]
Binder 1995	39	16	18	167	5-u	0.68 [0.55, 0.80]	0.91 [0.86, 0.95]
Malvehy 2014	168	297	70	1143	5-u	0.71 [0.64, 0.76]	0.79 [0.77, 0.81]
Hauschild 2014	47	30	18	35	5-u	0.72 [0.60, 0.83]	0.54 [0.41, 0.66]
Carli 2002b	8	6	3	36	5-u	0.73 [0.39, 0.94]	0.86 [0.71, 0.95]
Stanganelli 1999	8	3	3	16	5-u	0.73 [0.39, 0.94]	0.84 [0.60, 0.97]
Stanganelli 2005	23	21	8	82	4-c	0.74 [0.55, 0.88]	0.80 [0.71, 0.87]
Carrera 2016	89	145	30	213	4-u	0.75 [0.66, 0.82]	0.59 [0.54, 0.65]
Rubegni 2016	35	19	10	31	5-u	0.78 [0.63, 0.89]	0.62 [0.47, 0.75]
Tenenhaus 2010	25	29	7	166	4-u	0.78 [0.60, 0.91]	0.85 [0.79, 0.90]
Dolianitis 2005	16	4	4	16	5-u	0.80 [0.56, 0.94]	0.80 [0.56, 0.94]
Binder 1999	33	45	8	164	5-u	0.80 [0.65, 0.91]	0.78 [0.72, 0.84]
Seidenari 1998	25	3	6	56	5-c	0.81 [0.63, 0.93]	0.95 [0.86, 0.99]
Carli 2002	44	28	10	174	5-u	0.81 [0.69, 0.91]	0.86 [0.81, 0.91]
Carli 2003b	53	27	11	109	5*-u	0.83 [0.71, 0.91]	0.80 [0.72, 0.86]
Sboner 2004	35	20	7	90	5-u	0.83 [0.69, 0.93]	0.82 [0.73, 0.89]
Pagnanelli 2003	5	3	1	11	5-u	0.83 [0.36, 1.00]	0.79 [0.49, 0.95]
Carli 2003	37	71	7	85	5-u	0.84 [0.70, 0.93]	0.54 [0.46, 0.62]
Seidenari 2005	95	26	17	465	5-c	0.85 [0.77, 0.91]	0.95 [0.92, 0.97]
Annessi 2007	82	21	14	81	5*-u	0.85 [0.77, 0.92]	0.79 [0.70, 0.87]
Kittler 1998	20	7	3	20	5*-u	0.87 [0.66, 0.97]	0.74 [0.54, 0.89]
Gereli 2010	43	33	5	15	5-u	0.90 [0.77, 0.97]	0.31 [0.19, 0.46]
Pizzichetta 2004	35	3	4	66	5-u	0.90 [0.76, 0.97]	0.96 [0.88, 0.99]
Blum 2004b	76	208	8	545	4-u	0.90 [0.82, 0.96]	0.72 [0.69, 0.76]
Argenziano 1998	106	22	11	203	5-u	0.91 [0.84, 0.95]	0.90 [0.86, 0.94]
Guitera 2009b (Sydney)	40	55	4	32	5-c	0.91 [0.78, 0.97]	0.37 [0.27, 0.48]
Ferrari 2015	64	124	6	128	5*-u	0.91 [0.82, 0.97]	0.51 [0.44, 0.57]
Benelli 2001	11	3	1	35	5-u	0.92 [0.62, 1.00]	0.92 [0.79, 0.98]
Unlu 2014	22	36	2	55	5-u	0.92 [0.73, 0.99]	0.60 [0.50, 0.71]
Glud 2009	11	13	1	58	5-c	0.92 [0.62, 1.00]	0.82 [0.71, 0.90]
Pupelli 2013	22	33	2	39	5*-c	0.92 [0.73, 0.99]	0.54 [0.42, 0.66]
Zalaudek 2006	24	30	2	94	5-u	0.92 [0.75, 0.99]	0.76 [0.67, 0.83]
Piccolo 2002	12	2	1	326	5-u	0.92 [0.64, 1.00]	0.99 [0.98, 1.00]
Menzies 2005	12	27	1	38	4-u	0.92 [0.64, 1.00]	0.58 [0.46, 0.71]
Tan 2009	14	4	1	11	5-u	0.93 [0.68, 1.00]	0.73 [0.45, 0.92]
Alarcon 2014	86	128	6	123	5*-c	0.93 [0.86, 0.98]	0.49 [0.43, 0.55]
Gilmore 2010	34	17	2	16	5-u	0.94 [0.81, 0.99]	0.48 [0.31, 0.66]
Dal Pozzo 1999	159	79	9	466	5-u	0.95 [0.90, 0.98]	0.86 [0.82, 0.88]
Binder 1994	38	6	2	54	5-u	0.95 [0.83, 0.99]	0.90 [0.79, 0.96]
Dummer 1993	22	10	1	738	5*-c	0.96 [0.78, 1.00]	0.99 [0.98, 0.99]
Rubegni 2012	24	6	1	76	5*-u	0.96 [0.80, 1.00]	0.93 [0.85, 0.97]
Stolz 1994	47	3	1	28	5*-u	0.98 [0.89, 1.00]	0.90 [0.74, 0.98]
Piccolo 2014	33	69	0	63	5-u	1.00 [0.89, 1.00]	0.48 [0.39, 0.57]



Caption

Image-based evaluations of the accuracy of dermoscopy grouped by pathway categorisation for detecting for detecting invasive melanoma or melanocytic intraepidermal variants

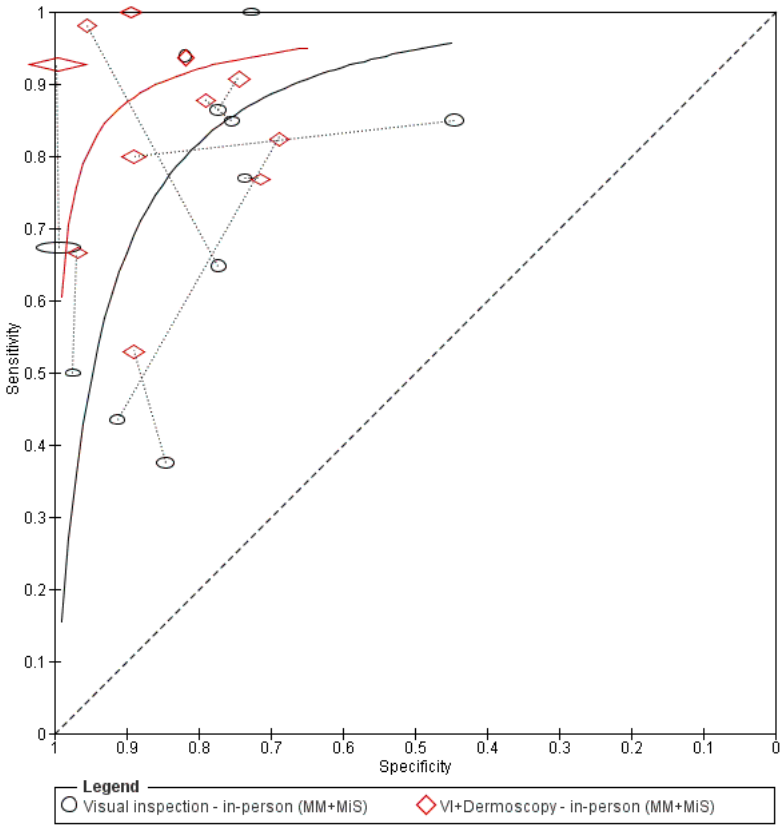
Figure 11 (Analysis 3)



Caption

Comparison of the accuracy of visual inspection with visual inspection plus dermoscopy for detection of invasive melanoma or melanocytic intraepidermal variants (MM+MiS) from in-person studies

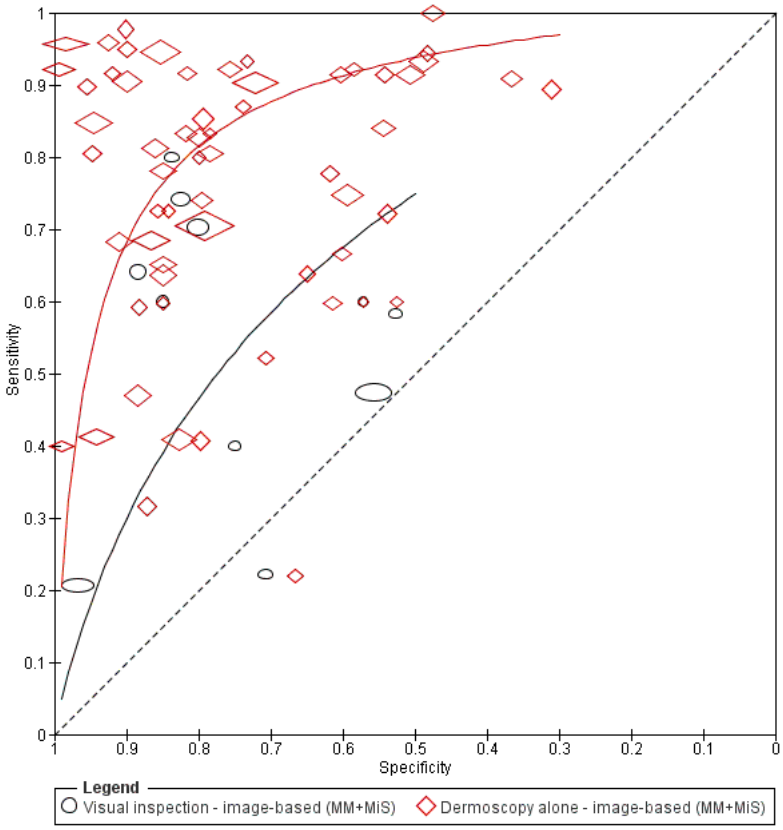
Figure 12 (Analysis 4)



Caption

Paired comparisons of the accuracy of visual inspection with visual inspection plus dermoscopy for detection of invasive melanoma or melanocytic intraepidermal variants (MM+MiS) from in-person studies

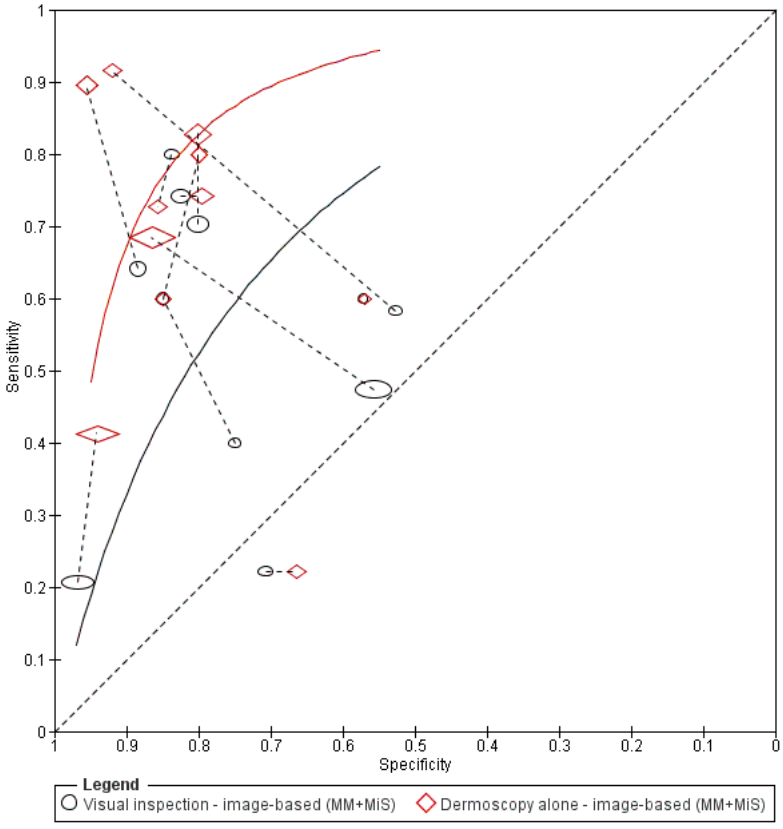
Figure 13 (Analysis 5)



Caption

Comparison of the accuracy of visual inspection with dermoscopy for detection of invasive melanoma or melanocytic intraepidermal variants (MM+MiS) from image-based studies

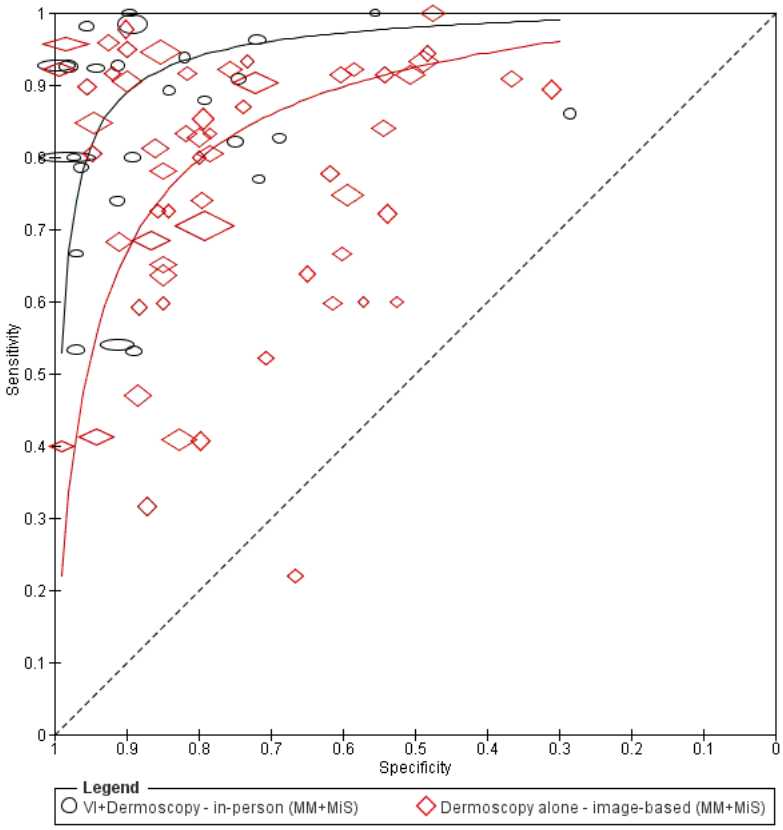
Figure 14 (Analysis 6)



Caption

Paired comparison of the accuracy of visual inspection versus dermoscopy for detection of invasive melanoma or melanocytic intraepidermal variants (MM+MiS) from paired image-based studies

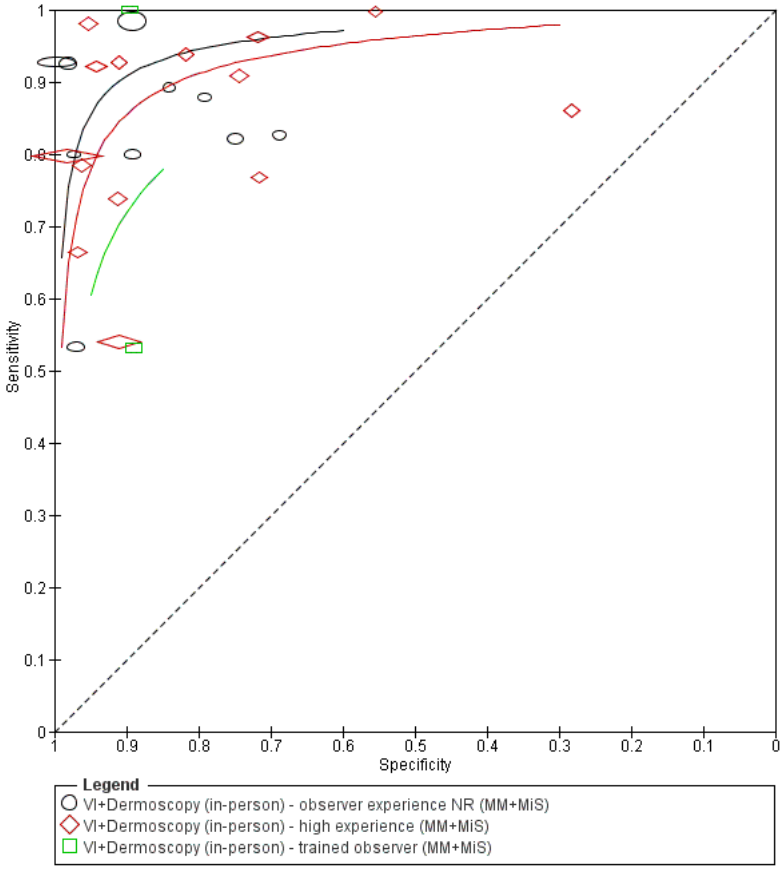
Figure 15 (Analysis 7)



Caption

Comparison of the accuracy of dermoscopy for detection of invasive melanoma or melanocytic intraepidermal variants (MM+MiS) between in-person (visual inspection plus dermoscopy) and image-based (dermoscopy) studies

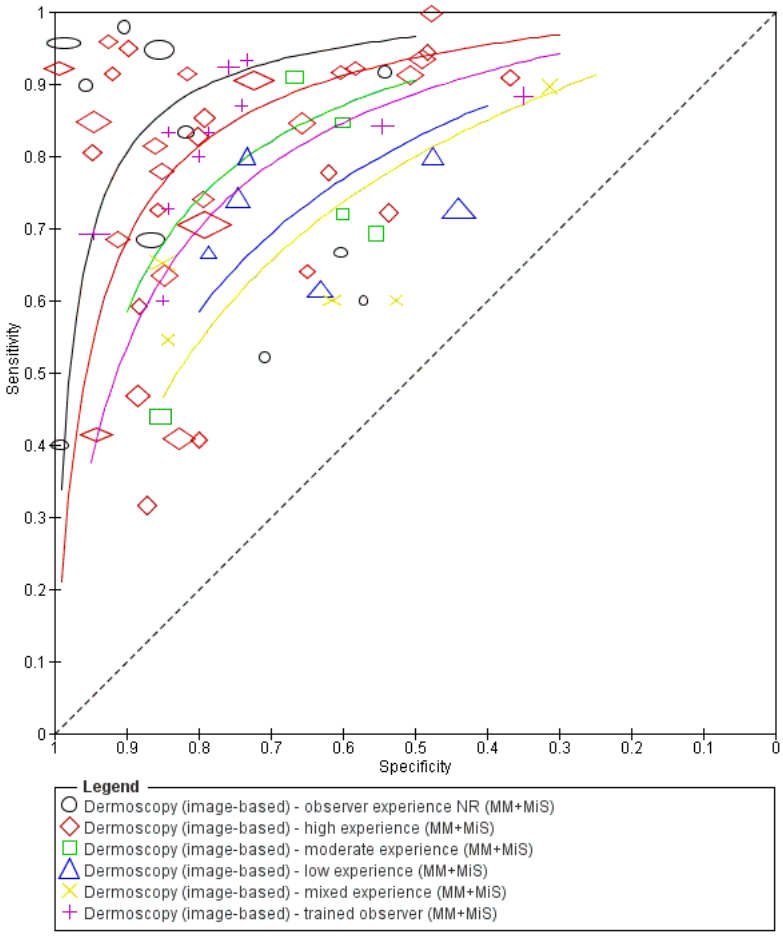
Figure 16 (Analysis 8)



Caption

Comparison of the accuracy of visual inspection plus dermoscopy for detection of invasive melanoma or melanocytic intraepidermal variants (MM+MiS) from in-person studies according to reported observer experience

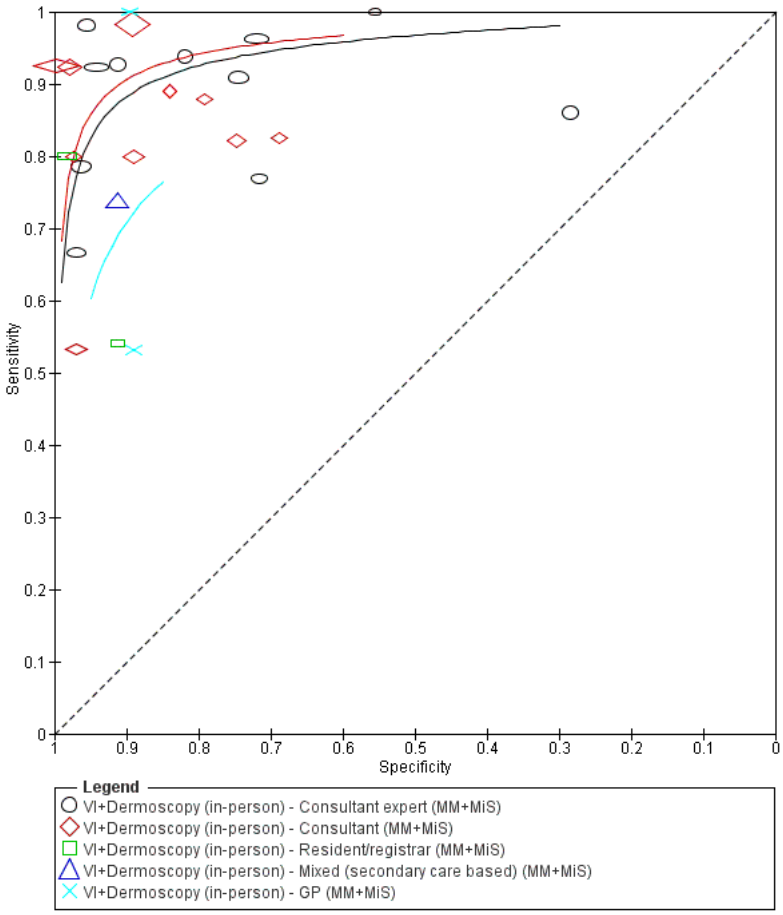
Figure 17 (Analysis 9)



Caption

Comparison of the accuracy of dermoscopy for detection of invasive melanoma or melanocytic intraepidermal variants (MM+MiS) from image-based studies according to observer experience

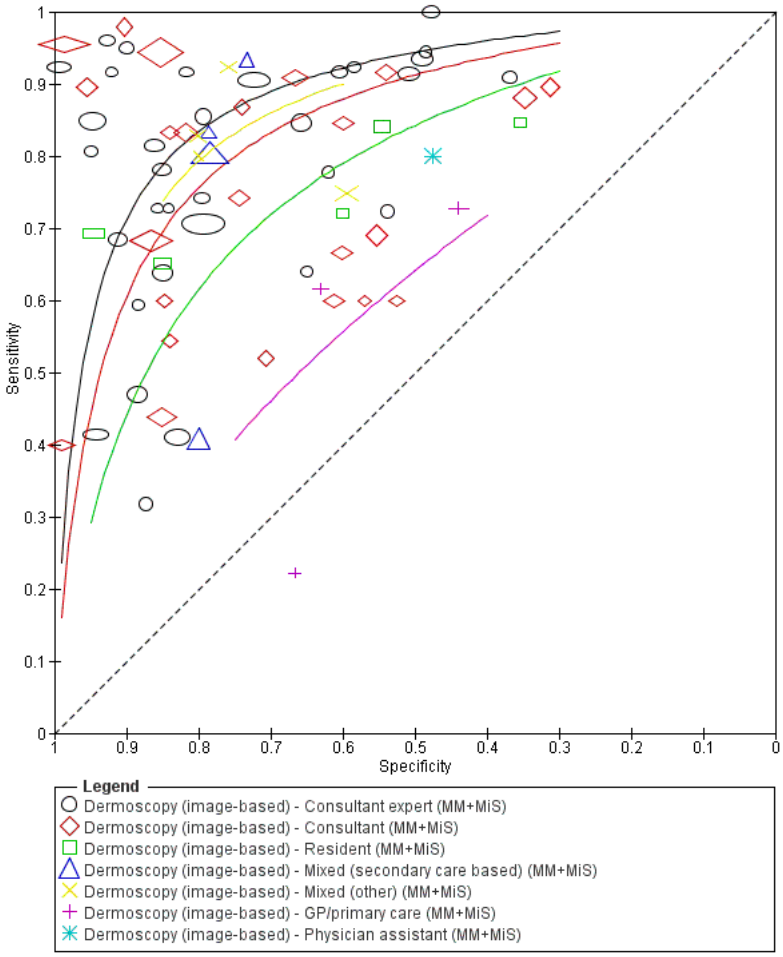
Figure 18 (Analysis 10)



Caption

Comparison of the accuracy of visual inspection plus dermoscopy for detection of invasive melanoma or melanocytic intraepidermal variants (MM+MiS) from in-person studies according to observer qualifications (summary ROC curves were not estimable from the model for Resident/Registrar and Mixed (secondary care based) experience groups)

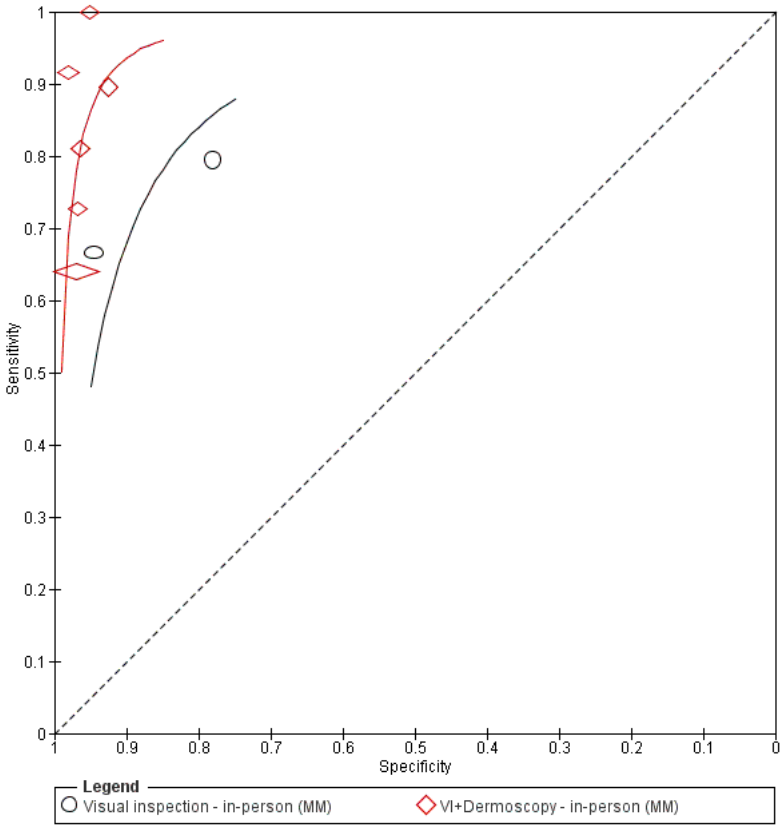
Figure 19 (Analysis 11)



Caption

Comparison of the accuracy of dermoscopy for detection of invasive melanoma or melanocytic intraepidermal variants (MM+MiS) from image-based studies according to observer qualification. (HSROC curves could not be estimated for Mixed (secondary care based) and Physician assistant groups).

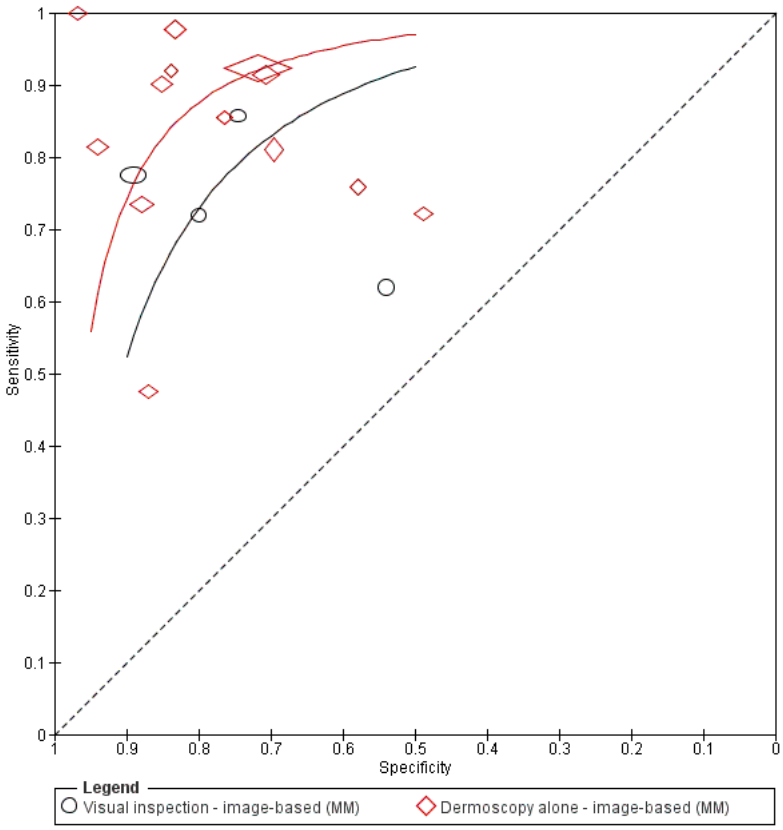
Figure 20 (Analysis 12)



Caption

Comparison of the accuracy of visual inspection with visual inspection plus dermoscopy for detection of invasive melanoma (MM) from in-person studies

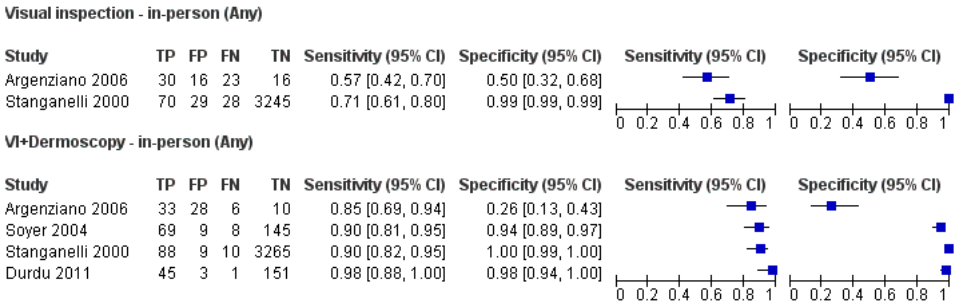
Figure 21 (Analysis 13)



Caption

Comparison of the accuracy of visual inspection with dermoscopy for detection of invasive melanoma (MM) from image-based studies

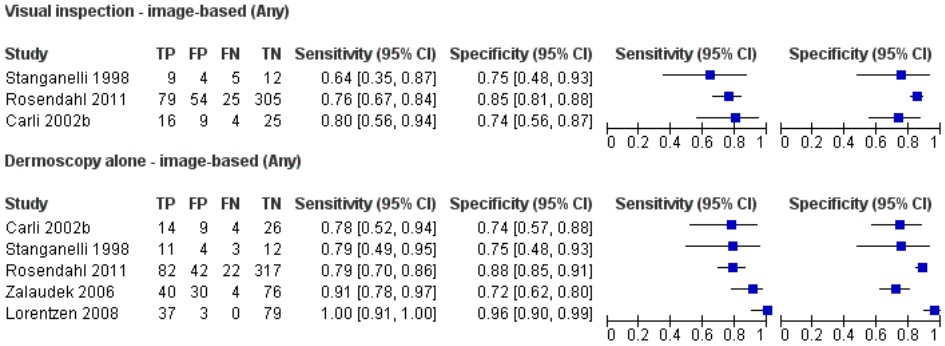
Figure 22 (Analysis 14)



Caption

Forest plot of tests: 9 Visual inspection - in-person (Any), 10 VI+Dermoscopy - in-person (Any).

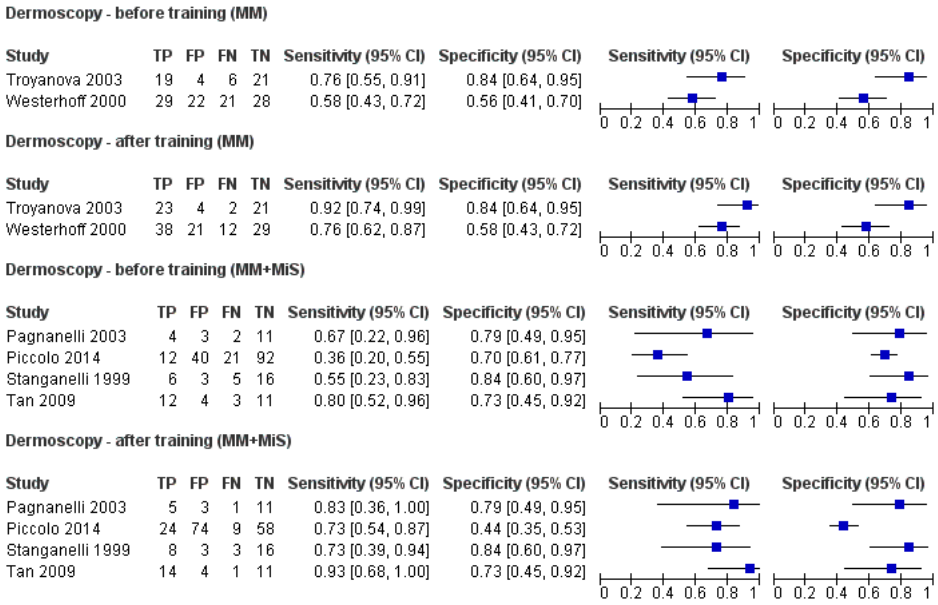
Figure 23 (Analysis 15)



Caption

Forest plot of tests: 11 Visual inspection - image-based (Any), 12 Dermoscopy alone - image-based (Any).

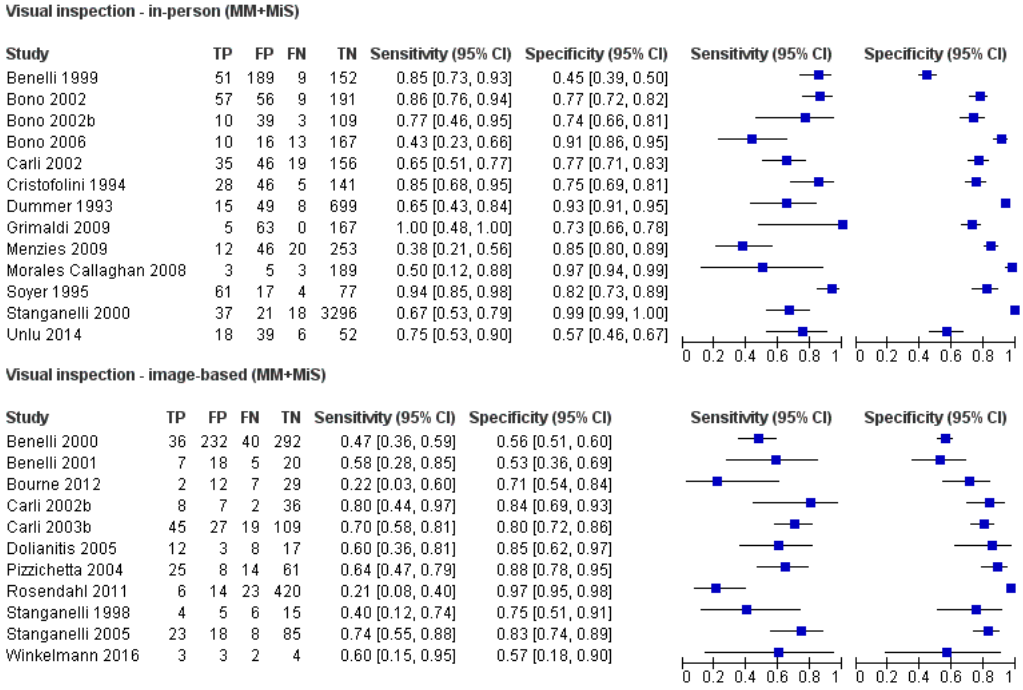
Figure 24 (Analysis 17)



Caption

Forest plot of tests: Accuracy of dermoscopy before and after dermoscopy training (MM and MM+MiS)

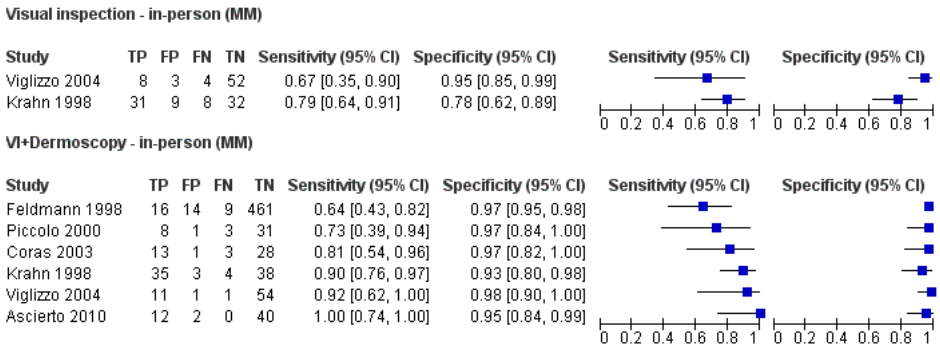
Figure 25 (Analysis 18)



Caption

Forest plot of tests: 5 Visual inspection - in-person (MM+MiS), 7 Visual inspection - image-based (MM+MiS).

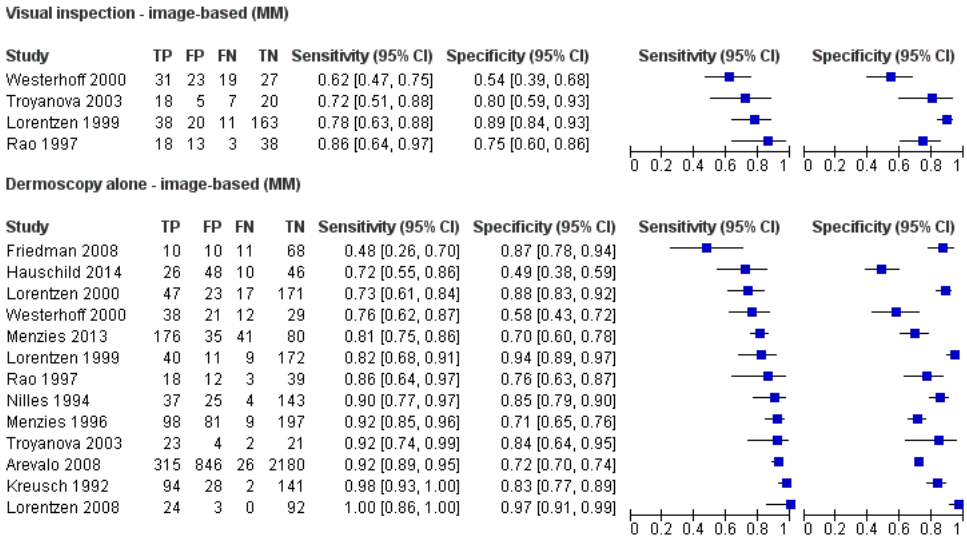
Figure 26 (Analysis 12)



Caption

Forest plot of tests: 1 Visual inspection - in-person (MM), 2 VI+Dermoscopy - in-person (MM).

Figure 27 (Analysis 13)



Caption

Forest plot of tests: 3 Visual inspection - image-based (MM), 4 Dermoscopy alone - image-based (MM).

Sources of support

Internal sources

- No sources of support provided

External sources

- NIHR Systematic Review Programme, UK
- The National Institute for Health Research (NIHR), UK
  The NIHR, UK, is the largest single funder of the Cochrane Skin Group

Feedback

Appendices

1 Current content and structure of the Programme Grant

List of reviews	Estimated number of studies
Diagnosis of melanoma	
1. Visual inspection versus visual inspection plus dermoscopy	120
2. Teledermatology	12
3. Mobile phone applications	2
4. Computer-aided diagnosis: dermoscopy based and spectroscopy based techniques	37
5. Reflectance confocal microscopy	19
6. High frequency ultrasound	3
7. Overview: comparing the accuracy of tests for which sufficient evidence was identified either alone or in combination	–
Diagnosis of keratinocyte skin cancer (basal cell carcinoma and cutaneous squamous cell carcinoma)	
8. Visual inspection ± dermoscopy	22
9. Computer aided diagnosis: dermoscopy based and spectroscopy based techniques	3
10. Optical coherence tomography	6
11. Reflectance confocal microscopy	9
12. High frequency ultrasound	1
13. Exfoliative cytology	5
14. Overview: comparing the accuracy of tests for which sufficient evidence was identified either alone or in combination	–
Staging of melanoma	
15. Ultrasound	25 to 30
16. Computer tomography	5 to 10
17. Positron emission tomography or positron emission tomography-computer tomography	20 to 25
18. Magnetic resonance imaging	5
19. Sentinel lymph node biopsy ± high frequency ultrasound	70
20. Overview: comparing the accuracy of tests for which sufficient evidence was identified either alone or in combination	–
Staging of cutaneous squamous cell carcinoma	
21. Imaging tests review	10 to 15
22. Sentinel lymph node biopsy ± high frequency ultrasound	15 to 20

2 Content of algorithms used to assist melanoma diagnosis using dermoscopy

Pattern analysis	ABCD	ABCD (revised)	ABCDE	Seven-point checklist

Pehamberger 1987	Stolz 1994	Blum 2003	Kittler 1999	Argenziano 1998 Argenziano 2011 (revised)
§ irregular and multicomponent pigmentary network pattern, § peripheral dark network patches, § sharp network margin, § pseudopods, § radial streaming, § blue-grey areas, § pigment dots (blotches, black dots, brown globules), § black dots at periphery, § whitish veil, § depigmentation and hypopigmented areas, § erythema, § telangiectasia, § comedo-like openings, milia-like cysts, § red-blue areas.	§ Asymmetry score x 1.3 calculated according to the colours and structures present within the lesion and not only with respect to the contour of the lesion § + Border score x 0.1 for each of 8 lesion segments presenting with an abrupt cut-off of pigment pattern the score was increased by one point. Maximum border score 8 § + Colour score x 0.5 Up to 6 different colours counted: white, red, blue-gray light-brown, dork-brown, and black § + Differential structure score x 0.5 Five main structural features: homogeneous areas network, streaks, dots, and 'globules according to size. Thresholds >5.45 or >4.75	§ Asymmetry of the outer shape in at least 1 axis (+1); Plus - Asymmetry of the differential structures inside the lesion in at least 1 axis (+1); § Border - Abrupt cutoff of network at the border of the lesion in at least 1 quarter of the circumference (+1); § Color - Three or more colors (+1); § Differential structures - Three or more differential structures (+1) Threshold >=4	As for ABCD but with addition of 'E' for enlargement or change Patient self-report of change in lesion size, color, or shape within the last year or whether they experienced any sign of ulceration or spontaneous bleeding. De novo appearance of a lesion within the last year was regarded as change in size ABCD-E score calculated by adding 1.2 to the standard ABCD score for changing lesions and subtracting 0.8 from the standard ABCD score for nonchanging lesions 6 thresholds tested, no single one recommended	Major criteria: § atypical network § blue-white veil § atypical vascular pattern Minor criteria § irregular dots/globules, § irregular streaks, § irregular blotches § regression structures Major criteria score 2 points each; minor criteria score 1 point each. Threshold for excision >= 3. For the **revised seven-point checklist**, each criterion is given a score of 1 point, and the threshold for excision is >=1 point, rather than >=3 points.
Seven-point checklist **(for lesion FU)** Stanganelli 2015	**Three-point checklist** Soyer 2004	**Four-point checklist** di Meo 2016	**Risk stratification** Kenet 1994	**Risk stratification (modified)** Ascierto 1998 , Ascierto 2003 , Ascierto 2010
§ A score of 'no change' was assigned if all variables remained constant, with a tolerance of major axis change of 2 mm (Beer 2011; Terushkin 2012); § 'minor change' if there was only symmetrical change in structural or chromatic pattern; § 'moderate change' if either structural or chromatic changes were asymmetrical, but there were no melanoma-specific criteria; and 'major change' if there were asymmetrical structural and chromatic changes, or the appearance of melanoma-specific criteria (i.e. major or minor criteria on original seven-point checklist: blue-white veil, atypical or negative pigment network, atypical vascular patterns, irregular dots and globules, streaks, irregular blotches, peripheral pigmented structureless areas and regression.)	§ Asymmetry - in color and/or structure in one or two axes, § Atypical pigment network - pigmented network with thickened lines and irregular distribution, § Blue-white structures - any blue and/or white color within the lesion The presence of two or three criteria is suggestive for melanoma.	Same as three-point checklist but asymmetry given 2 points instead of 1 § Asymmetry of color and structure 1st axis 1 point § Asymmetry of color and structure 2nd axis 1 point § Irregular or thick pigmented network 1 point § Blue-white structure 1 point A total score >2 was used as cut-off	Stratum 1 (probable MM): § Pseudopods, § Radial streaming, § Heterogeneity of pigment network with thick dark extensions at the edge, § Blue-grey areas, § white scarlike areas and § presence of pigment network Stratum 2 (possible MM): § Marked irregular network with irregular pigment confluence, § Eccentricity of pigment network with darkest regions near edge	Very high risk § - pigment network and any classical ELM features specific for melanoma: § pseudopods, § radial streaming, § blue-gray veil, High risk § - pigment network and "subtle new ELM features that may suggest melanoma but often are also seen in atypical nevi", e.g. § Irregular brown globules at periphery § Irregular black dots at periphery § Hypopigmentation at lesion periphery
Menzies' checklist Menzies 1996	**Seven features for melanoma (7FFM)** Dal Pozzo 1999	**Chaos and clues** Rosendahl 2011	**CASH** Dolianitis 2005; Henning 2007; Henning 2008	
Negative features: § Point and axial symmetry of pigmentation § Presence of only a single colour Positive features: § Multiple (5-6) colors § Blue-white veil § Multiple brown dots § Multiple blue/gray § Peripheral black dots or globules § A broadened network § Pseudopods § Radial streaming § Scarlike depigmentation Threshold: both negative features absent and >=1 positive features present	Major features: § regression erythema, § radial streaming, § gray-blue veil, § irregularly distributed pseudopods; Minor features score 1 each: § unhomogeneity, § irregular pigment network, § sharp margin. Major features score 2 points each; minor features score 1 point each. Threshold: >=2	Chaos - asymmetry of color or structure (defined by basic principles of pattern analysis as revised by Kittler (2007).) Clues: § eccentric structureless zone (any color except skin color), § gray or blue structures, § peripheral black dots or clods, § segmental radial lines or pseudopods, § polymorphous vessels,	Color: light brown, dark brown, black, red, white, blue, each receive 1 point. Architectural disorder: nonuniformity of structures and their distribution in the lesion; benign melanocytic lesions having uniform structures and distribution. Absent/mild, moderate and marked architectural disorder receive 0, 1, and 2 points, respectively. Symmetry: § biaxial symmetry scores 0 § monoaxial symmetry scores 1 § biaxial asymmetry scores 2 Homogeneity/heterogeneity; 7 structures each score 1 § network, § dots/globules, § streaks/pseudopods,	

		§ white lines, thick reticular or branched lines, and parallel lines on ridges (acral lesions). Clues searched for in presence of chaos; both present for test positive	§ blue-whitish veil, § regression structures (blue-gray areas with or without peppering; scarring) § blotches (structureless regions of any color occupying >10% of the area of the lesion) § polymorphous blood vessels (dotted and irregular linear). A total CASH score (TCS) >=8 is suggestive for melanoma.	

3 Glossary of terms

Term	Definition
Atypical intraepidermal melanocytic variant	Unusual area of darker pigmentation contained within the epidermis that may progress to an invasive melanoma; includes melanoma *in situ* and lentigo maligna
Atypical naevi	Unusual looking but noncancerous mole or area of darker pigmentation of the skin
BRAF V600 mutation	BRAF is a human gene that makes a protein called B-Raf which is involved in the control of cell growth. BRAF mutations (damaged DNA) occur in around 40% of melanomas, which can then be treated with particular drugs.
BRAF inhibitors	Therapeutic agents which inhibit the serine-threonine protein kinase BRAF mutated metastatic melanoma.
Breslow thickness	A scale for measuring the thickness of melanomas by the pathologist using a microscope, measured in mm from the top layer of skin to the bottom of the tumour.
Congenital naevi	A type of mole found on infants at birth
Dermoscopy	Whereby a handheld microscope is used to allow more detailed, magnified, examination of the skin compared to examination by the naked eye alone
False negative	An individual who is truly positive for a disease, but whom a diagnostic test classifies them as disease-free.
False positive	An individual who is truly disease-free, but whom a diagnostic test classifies them as having the disease.
Histopathology/Histology	The study of tissue, usually obtained by biopsy or excision, for example under a microscope.
Incidence	The number of new cases of a disease in a given time period.
Index test	A diagnostic test under evaluation in a primary study
Lentigo maligna	Unusual area of darker pigmentation contained within the epidermis which includes malignant cells but with no invasive growth. May progress to an invasive melanoma
Lymph node	Lymph nodes filter the lymphatic fluid (clear fluid containing white blood cells) that travels around the body to help fight disease; they are located throughout the body often in clusters (nodal basins).
Melanocytic naevus	An area of skin with darker pigmentation (or melanocytes) also referred to as ‘moles’
Meta-analysis	A form of statistical analysis used to synthesise results from a collection of individual studies.
Metastases/metastatic disease	Spread of cancer away from the primary site to somewhere else through the bloodstream or the lymphatic system.
Micrometastases	Micrometastases are metastases so small that they can only be seen under a microscope.
Mitotic rate	Microscopic evaluation of number of cells actively dividing in a tumour.
Morbidity	Detrimental effects on health.
Mortality	Either (1) the condition of being subject to death; or (2) the death rate, which reflects the number of deaths per unit of population in relation to any specific region, age group, disease, treatment or other classification, usually expressed as deaths per 100, 1000, 10,000 or 100,000 people.
Multidisciplinary team	A team with members from different healthcare professions and specialties (e.g. urology, oncology, pathology, radiology, and nursing). Cancer care in the National Health Service (NHS) uses this system to ensure that all relevant health professionals are engaged to discuss the best possible care for that patient.
Prevalence	The proportion of a population found to have a condition.
Prognostic factors/indicators	Specific characteristics of a cancer or the person who has it which might affect the patient’s prognosis.
Receiver operating characteristic (ROC) plot	A plot of the sensitivity and 1 minus the specificity of a test at the different possible thresholds for test positivity; represents the diagnostic capability of a test with a range of binary test results
Receiver operating characteristic (ROC) analysis	The analysis of a ROC plot of a test to select an optimal threshold for test positivity
Recurrence	Recurrence is when new cancer cells are detected following treatment. This can occur either at the site of the original tumour or at other sites in the body.
Reference Standard	A test or combination of tests used to establish the final or ‘true’ diagnosis of a patient in an evaluation of a diagnostic test
Reflectance confocal microscopy (RCM)	A microscopic technique using infrared light (either in a handheld device or a static unit) that can create images of the deeper layers of the skin
Sensitivity	In this context the term is used to mean the proportion of individuals with a disease who have that disease correctly identified by the study test
Specificity	The proportion of individuals without the disease of interest (in this case with benign skin lesions) who have that absence of disease correctly identified by the study test
Staging	Clinical description of the size and spread of a patient’s tumour, fitting into internationally agreed categories.
Subclinical (disease)	Disease that is usually asymptomatic and not easily observable, e.g. by clinical or physical examination.
Systemic treatment	Treatment, usually given by mouth or by injection, that reaches and affects cancer cells throughout the body rather than targeting one specific area.

4 Proposed sources of heterogeneity

i. Population characteristics

- general versus higher risk populations
- patient population: Primary /secondary / specialist unit
- lesion suspicion: general suspicion/atypical/equivocal/NR
- lesion type: any pigmented; melanocytic
- inclusion of multiple lesions per participant
- ethnicity

ii. Index test characteristics

- the nature of and definition of criteria for test positivity
- observer experience with the index test
- approaches to lesion preparation (e.g., the use of oil or antiseptic gel for dermoscopy)

iii. Reference standard characteristics

- reference standard used
- whether histology-reporting meets pathology-reporting guidelines
- use of excisional versus diagnostic biopsy
- whether two independent dermatopathologists reviewed histological diagnosis

iv. Study quality

- consecutive or random sample of participants recruited
- index test interpreted blinded to the reference standard result
- index test interpreted blinded to the result of any other index test
- presence of partial or differential verification bias (whereby only a sample of those subject to the index test are verified by the reference test or by the same reference test with selection dependent on the index test result)
- use of an adequate reference standard
- overall risk of bias

5 Final search strategies

Melanoma search strategies to August 2016

Database: Ovid MEDLINE(R) 1946 to August week 3 2016

Search strategy:

1 exp melanoma/

2 exp skin cancer/

3 exp basal cell carcinoma/

4 basalioma$1.ti,ab.

5 ((basal cell or skin) adj2 (cancer$1 or carcinoma$1 or mass or masses or tumour$1 or tumor$1 or neoplasm$1 or adenoma$1 or epithelioma$1 or lesion$1 or malignan$ or nodule$1)).ti,ab.

6 (pigmented adj2 (lesion$1 or mole$ or nevus or nevi or naevus or naevi or skin)).ti,ab.

7 (melanom$1 or nonmelanoma$1 or non-melanoma$1 or melanocyt$ or non-melanocyt$ or nonmelanocyt$ or keratinocyt$).ti,ab.

8 nmsc.ti,ab.

9 (squamous cell adj2 (cancer$1 or carcinoma$1 or mass or masses or tumor$1 or tumour$1 or neoplasm$1 or adenoma$1 or epithelioma$1 or epithelial or lesion$1 or malignan$ or nodule$1) adj2 (skin or epiderm$ or cutaneous)).ti,ab.

10 (BCC or CSCC or NMSC).ti,ab.

11 keratinocy$.ti,ab.

12 Keratinocytes/

13 or/1-12

14 dermoscop$.ti,ab.

15 dermatoscop$.ti,ab.

16 photomicrograph$.ti,ab.

17 exp epiluminescence microscopy/

18 (epiluminescence adj2 microscop$).ti,ab.

19 (confocal adj2 microscop$).ti,ab.

20 (incident light adj2 microscop$).ti,ab.

21 (surface adj2 microscop$).ti,ab.

22 (visual adj (inspect$ or examin$)).ti,ab.

23 ((clinical or physical) adj examin$).ti,ab.

24 3 point.ti,ab.

25 three point.ti,ab.

26 pattern analys$.ti,ab.

27 ABCD$.ti,ab.

28 menzies.ti,ab.

29 7 point.ti,ab.

30 seven point.ti,ab.

31 (digital adj2 (dermoscop$ or dermatoscop$)).ti,ab.

32 artificial intelligence.ti,ab.

33 AI.ti,ab.

34 computer assisted.ti,ab.

35 computer aided.ti,ab.

36 neural network$.ti,ab.

37 exp diagnosis, computer-assisted/

38 MoleMax.ti,ab.

39 image process$.ti,ab.

40 automatic classif$.ti,ab.

41 image analysis.ti,ab.

42 SIAscop$.ti,ab.

43 Aura.ti,ab.

44 (optical adj2 scan$).ti,ab.

45 MelaFind.ti,ab.

46 SIMSYS.ti,ab.

47 MoleMate.ti,ab.

48 SolarScan.ti,ab.

49 VivaScope.ti,ab.

50 (high adj3 ultraso$).ti,ab.
51 (canine adj2 detect$).ti,ab.
52 ((mobile or cell or cellular or smart) adj ((phone$1 adj2 app$1) or application$1)).ti,ab.
53 smartphone$.ti,ab.
54 (DermoScan or SkinVision or DermLink or SpotCheck).ti,ab.
55 Mole Detective.ti,ab.
56 Spot Check.ti,ab.
57 (mole$1 adj2 map$).ti,ab.
58 (total adj2 body).ti,ab.
59 exfoliative cytolog$.ti,ab.
60 digital analys$.ti,ab.
61 (image$1 adj3 software).ti,ab.
62 (teledermatolog$ or tele-dermatolog$ or telederm or tele-derm or teledermoscop$ or tele-dermoscop$ or teledermatoscop$ or tele-dermatoscop$).ti,ab.
63 (optical coherence adj (technolog$ or tomog$)).ti,ab.
64 (computer adj2 diagnos$).ti,ab.
65 exp sentinel lymph node biopsy/
66 (sentinel adj2 node).ti,ab.
67 nevisense.mp. or HFUS.ti,ab.
68 electrical impedance spectroscopy.ti,ab.
69 history taking.ti,ab.
70 patient history.ti,ab.
71 (naked eye adj (exam$ or assess$)).ti,ab.
72 (skin adj exam$).ti,ab.
73 physical examination/
74 ugly duckling.mp. or UD.ti,ab.
75 ((physician$ or clinical or physical) adj (exam$ or triage or recog$)).ti,ab.
76 ABCDE.mp. or VOC.ti,ab.
77 clinical accuracy.ti,ab.
78 Family Practice/ or Physicians, Family/ or clinical competence/
79 (confocal adj2 microscop$).ti,ab.
80 diagnostic algorithm$1.ti,ab.
81 checklist$.ti,ab.
82 virtual imag$1.ti,ab.
83 volatile organic compound$1.ti,ab.
84 dog$1.ti,ab.
85 gene expression analy$.ti,ab.
86 reflex transmission imag$.ti,ab.
87 thermal imaging.ti,ab.
88 elastography.ti,ab.
89 or/14-88
90 (CT or PET).ti,ab.
91 PET-CT.ti,ab.
92 (FDG or F18 or Fluorodeoxyglucose or radiopharmaceutical$).ti,ab.
93 exp Deoxyglucose/
94 deoxy-glucose.ti,ab.
95 deoxyglucose.ti,ab.
96 CATSCAN.ti,ab.
97 exp Tomography, Emission-Computed/
98 exp Tomography, X-ray computed/
99 positron emission tomograph$.ti,ab.
100 exp magnetic resonance imaging/
101 (MRI or fMRI or NMRI or scintigraph$).ti,ab.
102 exp echography/
103 Doppler echography.ti,ab.
104 sonograph$.ti,ab.
105 ultraso$.ti,ab.
106 doppler.ti,ab.
107 magnetic resonance imag$.ti,ab.
108 or/90-107
109 (stage$ or staging or metasta$ or recurrence or sensitivity or specificity or false negative$ or thickness$).ti,ab.
110 "Sensitivity and Specificity"/
111 exp cancer staging/
112 or/109-111
113 108 and 112
114 89 or 113
115 13 and 114

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations 29 August 2016

Search strategy:

1 basalioma$1.ti,ab.

2 ((basal cell or skin) adj2 (cancer$1 or carcinoma$1 or mass or masses or tumour$1 or tumor$1 or neoplasm$1 or adenoma$1 or epithelioma$1 or lesion$1 or malignan$ or nodule$1)).ti,ab.

3 (pigmented adj2 (lesion$1 or mole$ or nevus or nevi or naevus or naevi or skin)).ti,ab.

4 (melanom$1 or nonmelanoma$1 or non-melanoma$1 or melanocyt$ or non-melanocyt$ or nonmelanocyt$ or keratinocyt$).ti,ab.

5 nmsc.ti,ab.

6 (squamous cell adj2 (cancer$1 or carcinoma$1 or mass or masses or tumor$1 or tumour$1 or neoplasm$1 or adenoma$1 or epithelioma$1 or epithelial or lesion$1 or malignan$ or nodule$1) adj2 (skin or epiderm$ or cutaneous)).ti,ab.

7 (BCC or CSCC or NMSC).ti,ab.

8 keratinocy$.ti,ab.

9 or/1-8

10 dermoscop$.ti,ab.

11 dermatoscop$.ti,ab.

12 photomicrograph$.ti,ab.

13 (epiluminescence adj2 microscop$).ti,ab.

14 (confocal adj2 microscop$).ti,ab.

15 (incident light adj2 microscop$).ti,ab.

16 (surface adj2 microscop$).ti,ab.

17 (visual adj (inspect$ or examin$)).ti,ab.

18 ((clinical or physical) adj examin$).ti,ab.

19 3 point.ti,ab.

20 three point.ti,ab.

21 pattern analys$.ti,ab.

22 ABCD$.ti,ab.

23 menzies.ti,ab.

24 7 point.ti,ab.

25 seven point.ti,ab.

26 (digital adj2 (dermoscop$ or dermatoscop$)).ti,ab.

27 artificial intelligence.ti,ab.

28 AI.ti,ab.

29 computer assisted.ti,ab.

30 computer aided.ti,ab.

31 neural network$.ti,ab.

32 MoleMax.ti,ab.

33 image process$.ti,ab.

34 automatic classif$.ti,ab.

35 image analysis.ti,ab.

36 SIAscop$.ti,ab.

37 Aura.ti,ab.

38 (optical adj2 scan$).ti,ab.

39 MelaFind.ti,ab.

40 SIMSYS.ti,ab.

41 MoleMate.ti,ab.

42 SolarScan.ti,ab.

43 VivaScope.ti,ab.

44 (high adj3 ultraso$).ti,ab.

45 (canine adj2 detect$).ti,ab.

46 ((mobile or cell or cellular or smart) adj ((phone$1 adj2 app$1) or application$1)).ti,ab.

47 smartphone$.ti,ab.

48 (DermoScan or SkinVision or DermLink or SpotCheck).ti,ab.

49 Mole Detective.ti,ab.

50 Spot Check.ti,ab.

51 (mole$1 adj2 map$).ti,ab.

52 (total adj2 body).ti,ab.

53 exfoliative cytolog$.ti,ab.

54 digital analys$.ti,ab.

55 (image$1 adj3 software).ti,ab.

56 (teledermatolog$ or tele-dermatolog$ or telederm or tele-derm or teledermoscop$ or tele-dermoscop$ or teledermatoscop$ or tele-dermatoscop$).ti,ab.

57 (optical coherence adj (technolog$ or tomog$)).ti,ab.

58 (computer adj2 diagnos$).ti,ab.

59 (sentinel adj2 node).ti,ab.

60 nevisense.mp. or HFUS.ti,ab.

61 electrical impedance spectroscopy.ti,ab.

62 history taking.ti,ab.

63 patient history.ti,ab.

64 (naked eye adj (exam$ or assess$)).ti,ab.

65 (skin adj exam$).ti,ab.

66 ugly duckling.mp. or UD.ti,ab.

67 ((physician$ or clinical or physical) adj (exam$ or triage or recog$)).ti,ab.

68 ABCDE.mp. or VOC.ti,ab.

69 clinical accuracy.ti,ab.

70 (Family adj (Practice or Physicians)).ti,ab.

71 (confocal adj2 microscop$).ti,ab.

72 clinical competence.ti,ab.

73 diagnostic algorithm$1.ti,ab.

74 checklist$.ti,ab.

75 virtual imag$1.ti,ab.

76 volatile organic compound$1.ti,ab.

77 dog$1.ti,ab.

78 gene expression analy$.ti,ab.

79 reflex transmission imag$.ti,ab.

80 thermal imaging.ti,ab.

81 elastography.ti,ab.

82 or/10-81

83 (CT or PET).ti,ab.

84 PET-CT.ti,ab.

85 (FDG or F18 or Fluorodeoxyglucose or radiopharmaceutical$).ti,ab.

86 deoxy-glucose.ti,ab.

87 deoxyglucose.ti,ab.

88 CATSCAN.ti,ab.

89 positron emission tomograph$.ti,ab.

90 (MRI or fMRI or NMRI or scintigraph$).ti,ab.

91 Doppler echography.ti,ab.

92 sonograph$.ti,ab.

93 ultraso$.ti,ab.

94 doppler.ti,ab.

95 magnetic resonance imag$.ti,ab.

96 or/83-95

97 (stage$ or staging or metasta$ or recurrence or sensitivity or specificity or false negative$ or thickness$).ti,ab.

98 96 and 97

99 82 or 98

100 9 and 99

Database: Embase 1974 to 29 August 2016

Search strategy:

1 *melanoma/

2 *skin cancer/

3 *basal cell carcinoma/

4 basalioma$.ti,ab.

5 ((basal cell or skin) adj2 (cancer$1 or carcinoma$1 or mass or masses or tumour$1 or tumor$1 or neoplasm$ or adenoma$ or epithelioma$ or lesion$ or malignan$ or nodule$)).ti,ab.

6 (pigmented adj2 (lesion$1 or mole$ or nevus or nevi or naevus or naevi or skin)).ti,ab.

7 (melanom$1 or nonmelanoma$1 or non-melanoma$1 or melanocyt$ or non-melanocyt$ or nonmelanocyt$ or keratinocyt$).ti,ab.

8 nmsc.ti,ab.

9 (squamous cell adj2 (cancer$1 or carcinoma$1 or mass or tumor$1 or tumour$1 or neoplasm$1 or adenoma$1 or epithelioma$1 or epithelial or lesion$1 or malignan$ or nodule$1) adj2 (skin or epiderm$ or cutaneous)).ti,ab.

10 (BCC or cscc).mp. or NMSC.ti,ab.

11 keratinocyte.ti,ab.

12 keratinocy$.ti,ab.

13 or/1-12

14 dermoscop$.ti,ab.

15 dermatoscop$.ti,ab.

16 photomicrograph$.ti,ab.

17 *epiluminescence microscopy/

18 (epiluminescence adj2 microscop$).ti,ab.

19 (confocal adj2 microscop$).ti,ab.

20 (incident light adj2 microscop$).ti,ab.

21 (surface adj2 microscop$).ti,ab.

22 (visual adj (inspect$ or examin$)).ti,ab.

23 ((clinical or physical) adj examin$).ti,ab.

24 3 point.ti,ab.

25 three point.ti,ab.

26 pattern analys$.ti,ab.

27 ABCD$.ti,ab.

28 menzies.ti,ab.

29 7 point.ti,ab.

30 seven point.ti,ab.

31 (digital adj2 (dermoscop$ or dermatoscop$)).ti,ab.

32 artificial intelligence.ti,ab.

33 AI.ti,ab.
34 computer assisted.ti,ab.
35 computer aided.ti,ab.
36 neural network$.ti,ab.
37 MoleMax.ti,ab.
38 exp diagnosis, computer-assisted/
39 image process$.ti,ab.
40 automatic classif$.ti,ab.
41 image analysis.ti,ab.
42 SIAscop$.ti,ab.
43 (optical adj2 scan$).ti,ab.
44 Aura.ti,ab.
45 MelaFind.ti,ab.
46 SIMSYS.ti,ab.
47 MoleMate.ti,ab.
48 SolarScan.ti,ab.
49 VivaScope.ti,ab.
50 confocal microscop$.ti,ab.
51 (high adj3 ultraso$).ti,ab.
52 (canine adj2 detect$).ti,ab.
53 ((mobile or cell$ or cellular or smart) adj ((phone$1 adj2 app$1) or application$1)).ti,ab.
54 smartphone$.ti,ab.
55 (DermoScan or SkinVision or DermLink or SpotCheck).ti,ab.
56 Spot Check.ti,ab.
57 Mole Detective.ti,ab.
58 (mole$1 adj2 map$).ti,ab.
59 (total adj2 body).ti,ab.
60 exfoliative cytolog$.ti,ab.
61 digital analys$.ti,ab.
62 (image$1 adj3 software).ti,ab.
63 (optical coherence adj (technolog$ or tomog$)).ti,ab.
64 (teledermatolog$ or tele-dermatolog$ or telederm or tele-derm or teledermoscop$ or tele-dermoscop$ or teledermatoscop$).mp. or tele-dermatoscop$.ti,ab.
65 (computer adj2 diagnos$).ti,ab.
66 *sentinel lymph node biopsy/
67 (sentinel adj2 node).ti,ab.
68 nevisense.ti,ab.
69 HFUS.ti,ab.
70 electrical impedance spectroscopy.ti,ab.
71 history taking.ti,ab.
72 patient history.ti,ab.
73 (naked eye adj (exam$ or assess$)).ti,ab.
74 (skin adj exam$).ti,ab.
75 *physical examination/
76 ugly duckling.ti,ab.
77 UD sign$.ti,ab.
78 ((physician$ or clinical or physical) adj (exam$ or recog$ or triage)).ti,ab.
79 ABCDE.ti,ab.
80 clinical accuracy.ti,ab.
81 *general practice/
82 (confocal adj2 microscop$).ti,ab.
83 clinical competence/
84 diagnostic algorithm$.ti,ab.
85 checklist$1.ti,ab.
86 virtual image$1.ti,ab.
87 volatile organic compound$1.ti,ab.
88 VOC.ti,ab.
89 dog$1.ti,ab.
90 gene expression analys$.ti,ab.
91 reflex transmission imaging.ti,ab.
92 thermal imaging.ti,ab.
93 elastography.ti,ab.
94 dog$1.ti,ab.
95 gene expression analys$.ti,ab.
96 reflex transmission imaging.ti,ab.
97 thermal imaging.ti,ab.
98 elastography.ti,ab.
99 or/14-93
100 PET-CT.ti,ab.

101 (CT or PET).ti,ab.

102 (FDG or F18 or Fluorodeoxyglucose or radiopharmaceutical$).ti,ab.

103 exp Deoxyglucose/

104 CATSCAN.ti,ab.

105 deoxyglucose.ti,ab.

106 deoxy-glucose.ti,ab.

107 *positron emission tomography/

108 *computer assisted tomography/

109 positron emission tomograph$.ti,ab.

110 *nuclear magnetic resonance imaging/

111 (MRI or fMRI or NMRI or scintigraph$).ti,ab.

112 *echography/

113 Doppler.ti,ab.

114 sonograph$.ti,ab.

115 ultraso$.ti,ab.

116 magnetic resonance imag$.ti,ab.

117 or/100-116

118 (stage$ or staging or metasta$ or recurrence or sensitivity or specificity or false negative$ or thickness$).ti,ab.

119 "Sensitivity and Specificity"/

120 *cancer staging/

121 or/118-120

122 117 and 121

123 99 or 122

124 13 and 123

Database: Cochrane Library (Wiley) 2016 searched 30 August 2016 CDSR Issue 8 of 12 2016 CENTRAL Issue 7 of 12 2016 HTA Issue 3 of 4 July 2016 DARE Issue 3 of 4 2015

Search strategy:

#1 melanoma* or nonmelanoma* or non-melanoma* or melanocyt* or non-melanocyt* or nonmelanocyt* or keratinocyte*

#2 MeSH descriptor: [Melanoma] explode all trees

#3 "skin cancer*"

#4 MeSH descriptor: [Skin Neoplasms] explode all trees

#5 skin near/2 (cancer* or carcinoma* or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*)

#6 nmsc

#7 "squamous cell" near/2 (cancer* or carcinoma* or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*) near/2 (skin or epiderm* or cutaneous)

#8 "basal cell" near/2 (cancer* or carcinoma* or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*)

#9 pigmented near/2 (lesion* or nevus or mole* or naevi or naevus or nevi or skin)

#10 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9

#11 dermoscop*

#12 dermatoscop*

#13 Photomicrograph*

#14 MeSH descriptor: [Dermoscopy] explode all trees

#15 confocal near/2 microscop*

#16 epiluminescence near/2 microscop*

#17 incident next light near/2 microscop*

#18 surface near/2 microscop*

#19 "visual inspect*"

#20 "visual exam*"

#21 (clinical or physical) next (exam*)

#22 "3 point"

#23 "three point"

#24 "pattern analys*"

#25 ABDC

#26 menzies

#27 "7 point"

#28 "seven point"

#29 digital near/2 (dermoscop* or dermatoscop*)

#30 "artificial intelligence"

#31 "AI"

#32 "computer assisted"

#33 "computer aided"

#34 AI

#35 "neural network*"

#36 MoleMax

#37 "computer diagnosis"

#38 "image process*"

#39 "automatic classif*"

#40 SIAscope

#41 "image analysis"

#42 "optical near/2 scan*"

#43 Aura

#44 MelaFind

#45 SIMSYS

#46 MoleMate

#47 SolarScan

#48 Vivascope

#49 "confocal microscopy"

#50 high near/3 ultraso*

#51 canine near/2 detect*

#52 Mole* near/2 map*

#53 total near/2 body

#54 mobile* or smart near/2 phone*

#55 cell next phone*

#56 smartphone*

#57 "mitotic index"

#58 DermoScan or SkinVision or DermLink or SpotCheck

#59 "Mole Detective"

#60 "Spot Check"

#61 mole* near/2 map*

#62 total near/2 body

#63 "exfoliative cytolog*"

#64 "digital analys*"

#65 image near/3 software

#66 teledermatolog* or tele-dermatolog* or telederm or tele-derm or teledermoscop* or tele-dermoscop* or teledermatoscop* or tele-dermatolog*

#67 "optical coherence" next (technolog* or tomog*)

#68 computer near/2 diagnos*

#69 sentinel near/2 node*

#70 #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65 or #66 or #67 or #68 or #69

#71 ultraso*

#72 sonograph*

#73 MeSH descriptor: [Ultrasonography] explode all trees

#74 Doppler

#75 CT or PET or PET-CT

#76 "CAT SCAN" or "CATSCAN"

#77 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#78 MeSH descriptor: [Tomography, X-Ray Computed] explode all trees

#79 MRI

#80 MeSH descriptor: [Magnetic Resonance Imaging] explode all trees

#81 MRI or fMRI or NMRI or scintigraph*

#82 "magnetic resonance imag*"

#83 MeSH descriptor: [Deoxyglucose] explode all trees

#84 deoxyglucose or deoxy-glucose

#85 "positron emission tomograph*"

#86 #71 or #72 or #73 or #74 or #75 or #76 or #77 or #78 or #79 or #80 or #81 or #82 or #83 or #84 or #85

#87 stage* or staging or metasta* or recurrence or sensitivity or specificity or "false negative*" or thickness*

#88 MeSH descriptor: [Neoplasm Staging] explode all trees

#89 #87 or #88

#90 #89 and #86

#91 #70 or #90

#92 #10 and #91

#93 BCC or CSCC or NMCS

#94 keratinocy*

#95 #93 or #94

#96 #10 or #95

#97 nevisense

#98 HFUS

#99 "electrical impedance spectroscopy"

#100 "history taking"

#101 "patient history"

#102 naked next eye near/1 (exam* or assess*)

#103 skin next exam*

#104 "ugly duckling" or (UD sign*)

#105 MeSH descriptor: [Physical Examination] explode all trees

#106 (physician* or clinical or physical) near/1 (exam* or recog* or triage*)

#107 ABCDE

#108 "clinical accuracy"

#109 MeSH descriptor: [General Practice] explode all trees

#110 confocal near microscop*

#111 "diagnostic algorithm*"

#112 MeSH descriptor: [Clinical Competence] explode all trees

#113 checklist*

#114 "virtual image*"

#115 "volatile organic compound*"

#116 dog or dogs

#117 VOC

#118 "gene expression analys*"

#119 "reflex transmission imaging"

#120 "thermal imaging"

#121 elastography

#122 #97 or #98 or #99 or #100 or #101 or #102 or #103 or #104 or #105 or #106 or #107 or #108 or #109 or #110 or #111 or #112 or #113 or #114 or #115 or #116 or #117 or #118 or #119 or #120 or #121

#123 #70 or #122

#124 #96 and #123

#125 #96 and #90

#126 #125 or #124

#127 #10 and #126

Database : CINAHL Plus (EBSCO) 1937 to 30 August 2016

Search strategy:

S1 (MH "Melanoma") OR (MH "Nevi and Melanomas+")

S2 (MH "Skin Neoplasms+")

S3 (MH "Carcinoma, Basal Cell+")

S4 basalioma*

S5 (basal cell) N2 (cancer* or carcinoma* or mass or masses or tumor* or tumour* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*)

S6 (pigmented) N2 (lesion* or mole* or nevus or nevi or naevus or naevi or skin)

S7 melanom* or nonmelanoma* or non-melanoma* or melanocyt* or non-melanocyt* or nonmelanocyt*

S8 nmsc

S9 TX BCC or cscc or NMSC

S10 (MH "Keratinocytes")

S11 keratinocyt*

S12 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11

S13 dermoscop* or dermatoscop* or photomicrograph* or (3 point) or (three point) or ABCD* or menzies or (7 point) or (seven point) or AI or Molemax or SIASCOP* or Aura or MelaFind or SIMSYS or MoleMate or SolarScan or smartphone* or DermoScan or SkinVision or DermLink or SpotCheck

S14 (epiluminescence or confocal or incident or surface) N2 (microscop*)

S15 visual N1 (inspect* or examin*)

S16 (clinical or physical) N1 (examin*)

S17 pattern analys*

S18 (digital) N2 (dermoscop* or dermatoscop*)

S19 (artificial intelligence)

S20 (computer) N2 (assisted or aided)

S21 (neural network*)

S22 (MH "Diagnosis, Computer Assisted+")

S23 (image process*)

S24 (automatic classif*)

S25 (image analysis)

S26 SIAScop*

S27 (optical) N2 (scan*)

S28 (high) N3 (ultraso*)

S29 elastography

S30 (mobile or cell or cellular or smart) N2 (phone*) N2 (app or application*)

S31 (mole*) N2 (map*)

S32 total N2 body

S33 exfoliative cytolog*

S34 digital analys*

S35 image N3 software

S36 teledermatolog* or tele-dermatolog* or telederm or tele-derm or teledermoscop* or tele-dermoscop* or teledermatoscop* or tele-dermatoscop* teledermatolog* or tele-dermatolog* or telederm or tele-derm or teledermoscop*

S37 (optical coherence) N1 (technolog* or tomog*)

S38 computer N2 diagnos*

S39 sentinel N2 node

S40 (MH "Sentinel Lymph Node Biopsy")

S41 nevisense or HFUS or checklist* or VOC or dog*

S42 electrical impedance spectroscopy

S43 history taking

S44 "Patient history"

S45 naked eye

S46 skin exam*

S47 physical exam*

S48 ugly duckling

S49 UD sign*

S50 (physician* or clinical or physical) N1 (exam*)

S51 clinical accuracy

S52 general practice

S53 (physician* or clinical or physical) N1 (recog* or triage)

S54 confocal microscop*

S55 clinical competence

S56 diagnostic algorithm*

S57 checklist*

S58 virtual image*

S59 volatile organic compound*

S60 gene expression analys*

S61 reflex transmission imag*

S62 thermal imaging

S63 S13 or S14 or S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47 OR S48 OR S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62

S64 CT or PET

S65 PET-CT

S66 FDG or F18 or Fluorodeoxyglucose or radiopharmaceutical*

S67 (MH "Deoxyglucose+")

S68 deoxy-glucose or deoxyglucose

S69 CATSCAN

S70 CAT-SCAN

S71 (MH "Deoxyglucose+")

S72 (MH "Tomography, Emission-Computed+")

S73 (MH "Tomography, X-Ray Computed")

S74 positron emission tomograph*

S75 (MH "Magnetic Resonance Imaging+")

S76 MRI or fMRI or NMRI or scintigraph*

S77 echography

S78 doppler

S79 sonograph*

S80 ultraso*

S81 magnetic resonance imag*

S82 S64 OR S65 OR S66 OR S67 OR S68 OR S69 OR S70 OR S71 OR S72 OR S73 OR S74 OR S75 OR S76 OR S77 OR S78 OR S79 OR S80 OR S81

S83 stage* or staging or metasta* or recurrence or sensitivity or specificity or (false negative*) or thickness

S84 (MH "Neoplasm Staging")

S85 S83 OR S84

S86 S82 AND S85

S87 S63 OR S86

S88 S12 AND S87

Database: Science Citation Index SCI Expanded (Web of Science) 1900 to 30 August 2016

Conference Proceedings Citation Index (Web of Science) 1900 to 1 September 2016

Search strategy:

#1 (melanom* or nonmelanom* or non-melanoma* or melanocyt* or non-melanocyt* or nonmelanocyt* or keratinocyt*)

#2 (basalioma*)

#3 ((skin) near/2 (cancer* or carcinoma or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*))

#4 ((basal) near/2 (cancer* or carcinoma* or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*))

#5 ((pigmented) near/2 (lesion* or mole* or nevus or nevi or naevus or naevi or skin))

#6 (nmsc or BCC or NMSC or keratinocy*)

#7 ((squamous cell (cancer* or carcinoma* or mass or masses or tumour* or tumor* or neoplasm* or adenoma* or epithelioma* or lesion* or malignan* or nodule*))

#8 (skin or epiderm* or cutaneous)

#9 #8 AND #7

#10 #9 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1

#11 ((dermoscop* or dermatoscop* or photomicrograph* or epiluminescence or confocal or "incident light" or "surface microscop*" or "visual inspect*" or "physical exam*" or 3 point or three point or pattern analy* or ABCDE or menzies or 7 point or seven point or dermoscop* or dermatoscop* or AI or artificial or computer aided or computer assisted or neural network* or Molemax or image process* or automatic classif* or image analysis or siascope or optical scan* or Aura or melafind or simsys or molemate or solarscan or vivascope or confocal microscop* or high ultraso* or canine detect* or cellphone* or mobile* or phone* or smartphone or dermoscan or skinvision or dermlink or spotcheck or spot check or mole detective or mole map* or total body or exfoliative psychology or digital or image software or optical coherence or teledermatology or telederm* or teledermoscop* or teledermatoscop* or computer diagnos* or sentinel))

#12 ((nevisense or HFUS or impedance spectroscopy or history taking or patient history or naked eye or skin exam* or physical exam* or ugly duckling or UD sign* or physician* exam* or physical exam* or ABCDE or clinical accuracy or general practice or confocal microscop* or clinical competence or diagnostic algorithm* or checklist* or virtual image* or volatile organic or VOC or dog* or gene expression or reflex transmission or thermal imag* or elastography))

#13 #11 or #12

#14 ((PET or CT or FDG or deoxyglucose or deoxy-glucose or fluorodeoxy* or radiopharma* or CATSCAN or positron emission or computer assisted or nuclear magnetic or MRI or FMRI or NMRI or scintigraph* or echograph* or Doppler or sonograph* or ultraso* or magnetic reson*))

#15 ((stage* or staging or metast* or recurrence or sensitivity or specificity or false negative* or thickness*))

#16 #14 AND #15

#17 #16 OR #13

#18 #10 AND #17

Refined by: DOCUMENT TYPES: (MEETING ABSTRACT OR PROCEEDINGS PAPER)

6 Full text inclusion criteria

Criterion	Inclusion	Exclusion

Study design	**For diagnostic and staging reviews**
	• Any study for which a 2×2 contingency table can be extracted, e.g.
	  ◦ diagnostic case control studies
	  ◦ 'cross-sectional' test accuracy study with retrospective or prospective data collection
	  ◦ studies where estimation of test accuracy was not the primary objective but test results for both index and reference standard were available
	  ◦ RCTs of tests or testing strategies where participants were randomised between index tests and all undergo a reference standard (i.e. accuracy RCTs)
		• < 5 melanoma cases (diagnosis reviews)
		• < 10 participants (staging reviews)
		• Studies developing new criteria for diagnosis unless a separate 'test set' of images were used to evaluate the criteria (mainly digital dermoscopy)
		• Studies using 'normal' skin as controls
		• Letters, editorials, comment papers, narrative reviews
		• Insufficient data to construct a 2×2 table

Target condition	• Melanoma
	• Keratinocyte skin cancer (or non-melanoma skin cancer)
	  ◦ BCC or epithelioma
	  ◦ cSCC
		• Studies exclusively conducted in children
		• Studies of non-cutaneous melanoma or SCC

Population	**For diagnostic reviews** • Adults with a skin lesion suspicious for melanoma, BCC, or cSCC (other terms include pigmented skin lesion/nevi, melanocytic, keratinocyte, etc.) • Adults at high risk of developing melanoma skin cancer, BCC, or cSCC **For staging reviews** • Adults with a diagnosis of melanoma or cSCC undergoing tests for staging of lymph nodes or distant metastases or both	• People suspected of other forms of skin cancer • Studies conducted exclusively in children
Index tests	**For diagnosis** • Visual inspection/clinical examination • Dermoscopy/dermatoscopy • Teledermoscpoy • Smartphone/mobile phone applications • Digital dermoscopy/artificial intelligence • Confocal microscopy • Ocular coherence tomography • Exfoliative cytology • High frequency ultrasound • Canine odour detection • DNA expression analysis/gene chip analysis • Other **For staging** • CT • PET • PET-CT • MRI • Ultrasound +/fine needle aspiration cytology FNAC • SLNB +/high frequency ultrasound • Other Any test combination and in any order Any test positivity threshold Any variation in testing procedure (e.g. radioisotope used)	• Sentinel lymph biopsy for therapeutic rather than staging purposes • Tests to determine melanoma thickness • Tests to determine surgical margins/lesion borders • Tests to improve histopathology diagnose • LND
Reference standard	**For diagnostic studies** • Histopathology of the excised lesion • Clinical follow-up of non-excised/benign appearing lesions with later histopathology if suspicious • Expert diagnosis (studies should not be included if expert diagnosis is the sole reference standard) **For studies of imaging tests for staging** • Histopathology (via LND or SLMB) • Clinical/radiological follow-up • A combination of the above **For studies of SLNB accuracy for staging** • LND of both SLN+ and SLn participants to identify all diseased nodes • LND of SLN+ participants and follow-up of SLN participants to identify a subsequent nodal recurrence in a *previously investigated* nodal basin	**For diagnostic studies** • Exclude if any disease positive participants have diagnosis unconfirmed by histology • Exclude if > 50% of disease negative participants have diagnosis confirmed by expert opinion with no histology or follow-up • Exclude studies of referral accuracy, i.e. comparing referral decision with expert diagnosis, unless evaluations of teledermatology or mobile phone applications
BCC: basal cell carcinoma; cSCC: cutaneous squamous cell carcinoma; CT: computed tomography; FNAC: fine needle aspiration cytology; LND: lymph node dissection; MRI: magnetic resonance imaging; PET: positron emission tomography; PET-CT: positron emission tomography computed tomography; RCT: randomised controlled trial; SCC: squamous cell carcinoma; SLN+: positive sentinel lymph node; SLn: negative sentinel lymph node; SLNB: sentinel lymph node biopsy.		

7 Quality assessment (based on QUADAS-2)

Patient selection domain (1)

Selective recruitment of study participants can be a key influence on test accuracy. In general terms, all participants eligible to undergo a test should be included in a study, allowing for the intended use of that test within the context of the study. We considered studies that separately sampled malignant and benign lesions to have used a case-control design; and those that supplemented a series of suspicious lesions with additional malignant or benign lesions to be at unclear risk of bias

In terms of exclusions, we considered studies that excluded particular lesion types (e.g. lentigo maligna), particular lesion sites, or excluded lesions with lack of observer agreement (e.g. on histopathology) to be at high risk of bias. For image-based evaluations, some studies excluded lesions on the basis of image quality as an *a priori* exclusion criterion while others excluded lesions with inadequate images *post hoc*. In order to judge studies consistently, we considered all exclusions due to image quality in the Flow and Timing domain (Were all participants included in the analysis?).

In judging the applicability of patient populations to the review question, we considered restriction to particular lesion populations, such as melanocytic, nodular, high risk or restrictions by size to be of high concern for applicability. Studies that included only lesions with histopathology results were also considered of high concern for applicability on the basis that in usual practice, whether in primary, secondary or specialist care, a greater or lesser proportion of patients will have lesions with low levels of suspicion of malignancy such that they can be reassured and discharged, or followed up over a period of time. The restriction of a study sample to those with lesions undergoing biopsy or excision will therefore not adequately reflect a usual care setting. Furthermore, due to the invasive nature of sampling lesions for histology, studies are not likely to mandate biopsy or excision as a study requirement regardless of the index of suspicion (in which case restriction to those with histology would not be of concern in terms of having a representative population).

Given that diagnosis of skin cancer is primarily lesion-based, there is the potential for study participants with multiple lesions to contribute disproportionately to estimates of test accuracy, especially if they are at particular risk of having skin cancer. We considered studies that include a high number of lesions in relation to the number of study to be less representative than studies conducted in a more general population participants (i.e., if the difference between the number of included lesions and number of included participants is greater than 5%).

Index test domain (2)

Given the potential for subjective differences in test interpretation for melanoma, the interpretation of the index test blinded to the result of the reference standard is a key means of reducing bias. For prospective studies and retrospective studies that used the original index test interpretation, the diagnosis will by nature be interpreted and recorded before the result of the reference standard is known; however, studies using previously acquired images could be particularly susceptible to information bias. For these studies to be at low risk of bias, we required a clear indication that observers were unaware of the reference standard diagnosis at time of test interpretation. An item was also added to assess the presence of blinding between interpretations of different algorithms, however this item was not included in the overall assessment of risk of bias.

Pre-specification of the index test threshold was considered present if the study clearly reported that the threshold used was not data driven, i.e., was not based on study results. Studies that did not clearly describe the threshold used but that required clinicians to record a diagnosis or management decision for a lesion were considered to be unclear on this criterion. Studies reporting accuracy for multiple numeric thresholds, where ROC analysis was used to select the threshold, or that reported accuracy for the presence of independently significant lesion characteristics with no separate test set of lesions were considered at high risk of bias.

In terms of applicability of the index test to the review question, we required the test to be applied and interpreted as it would be in a clinical practice setting, i.e., in-person or face-to-face with the study participant, and by a single observer as opposed to a consensus decision or average across multiple observers. Image-based studies were considered to be of high concern.

Despite the often subjective nature of test interpretation, it is also important for study authors to outline the particular lesion characteristics that were considered to be indicative for melanoma, particularly where established algorithms or checklists were not used. Studies were considered of low concern if the threshold used was established in a prior study or sufficient threshold details were presented to allow replication.

The experience of the examiner will also impact on the applicability of study results. We required studies to describe the test interpreter as ‘experienced’ or ‘expert’ to have low concern about applicability.

Reference standard domain (3)

In an ideal study, consecutively recruited participants should all undergo incisional or excisional biopsy of the skin lesion regardless of level of clinical suspicion of melanoma. In reality, both partial and differential verification bias are likely. Partial verification bias may occur where histology is the only reference standard used, and only those participants with a certain degree of suspicion of malignancy based on the result of the index test undergo verification, the others either being excluded from the study or defined as being disease-negative without further assessment or follow-up, as discussed under ‘**Patient selection domain’**.

Differential verification bias will be present where other reference standards are used in addition to histological verification of suspicious lesions. A typical example of verification bias in skin cancer occurs when investigators do not biopsy people with benign-appearing lesions but instead follow them up for a period of time to determine whether any malignancy subsequently develops (these would be false-negatives on the index test). We defined an 'adequate' reference standard as: all disease-positive individuals having a histological reference standard either at the time of application of the index test or after a period of clinical follow-up; and at least 80% of disease-negative participants have received a histological diagnosis, with up to 20% undergoing at least three months' follow-up of benign-appearing lesions.

A further challenge is the potential for incorporation bias, i.e., where the result of the index test is used to help determine the reference standard diagnosis. It is normal practice for the clinical diagnosis (usually by visual inspection or dermoscopy) to be included on pathology request forms and for the histopathologist to use this diagnosis to help with the pathology interpretation. Although inclusion of such clinical information on the histopathology request form is theoretically a form of incorporation bias, blinded interpretation of the histopathology reference standard is not normal practice, and enforcement of such conditions would significantly limit the generalisability of the study results. Blinding to the index test (visual inspection or clinical diagnosis) was therefore recorded but did not contribute to our overall assessment of risk of bias for the reference standard domain.

In judging the applicability of the reference standard to our review question, scored studies as high concern around applicability if they used expert diagnosis (with no follow-up) as a reference standard in any patient, or did not report histology interpretation by a dermatopathologist.

Flow and timing domain (4)

In the ideal study, the diagnosis based on the index test and reference standard should be made consecutively or as near to each other in time as possible to avoid changes in lesion over time. For lesions with a histological reference standard, we have defined a one-month period as an appropriate interval between application of the index test and the reference standard. Studies reporting biopsy or excision ‘following’, ‘after’ or ‘subsequent to’ the visual inspection diagnosis (or using similar descriptors) were considered to have met this criterion. For studies using clinical follow-up, a minimum three-month follow-up period has been defined as at low risk of bias for detecting false-negatives. This interval was chosen based on a study showing that most false-negative melanomas will be diagnosed within three months of the initial negative index test although a small number will be diagnosed up to 12 months subsequently (Altamura 2008).

In assessing whether all patients were included in the analysis, we considered studies at high risk of bias if participants were excluded following recruitment. As discussed in the ‘Patient selection domain’, a priori exclusion of images on the basis of image quality was also considered under this item.

The following tables use text that was originally published in the QUADAS-2 tool by Whiting and colleagues (Whiting 2011 *).*

Item	Response (delete as required)
PARTICIPANT SELECTION (1) RISK OF BIAS	
1) Was a consecutive or random sample of participants or images enrolled?	**Yes** – if paper states consecutive or random **No** – if paper describes other method of sampling **Unclear** – if participant sampling not described
2) Was a case-control design avoided?	**Yes** – if consecutive or random or case-control design clearly not used **No** – if study described as case-control or describes sampling specific numbers of participants with particular diagnoses **Unclear** – if not described
3) Did the study avoid inappropriate exclusions, e.g., • 'difficult to diagnose' lesions not excluded • lesions not excluded on basis of disagreement between evaluators	**Yes** - if inappropriate exclusions were avoided **No** – if lesions were excluded that might affect test accuracy, e.g., 'difficult to diagnose' lesions, or where disagreement between evaluators was observed **Unclear** – if not clearly reported but there is suspicion that difficult to diagnose lesions may have been excluded
4) For between-person comparative studies only (i.e., allocating different tests to different study participants): • **A)** were the same participant selection criteria used for those allocated to each test? • **B)** was the potential for biased allocation between tests avoided through adequate generation of a randomised sequence? • **C)** was the potential for biased allocation between tests avoided through concealment of allocation prior to assignment?	**For A)** • **Yes** – if same selection criteria were used for each index test, **No** – if different selection criteria were used for each index test, **Unclear** – if selection criteria per test were not described, **N/A** – if only 1 index test was evaluated or all participants received all tests **For B)** • **Yes** – if adequate randomisation procedures are described, **No** – if inadequate randomisation procedures are described, **Unclear** – if the method of allocation to groups is not described (a description of 'random' or 'randomised' is insufficient), **N/A** – if only 1 index test was evaluated or all participants received all tests **For C)** • **Yes** – if appropriate methods of allocation concealment are described, **No** – if appropriate methods of allocation concealment are not described, **Unclear** – if the method of allocation concealment is not described (sufficient detail to allow a definite judgement is required), **N/A** – if only 1 index test was evaluated
Could the selection of participants have introduced bias? **For non-comparative and within-person-comparative studies** 1. If answers to all of questions 1), 2), and 3) 'Yes': 2. If answers to any 1 of questions 1), 2), or 3) 'No': 3. If answers to any 1 of questions 1), 2), or 3) 'Unclear':	**For non-comparative and within-person-comparative studies** 1. Risk is low 2. Risk is high 3. Risk unclear **For between-person comparative studies**

For between-person comparative studies 1. If answers to all of questions 1), 2), 3), and 4) 'Yes': 2. If answers to any 1 of questions 1), 2), 3), or 4) 'No': 3. If answers to any 1 of questions 1), 2), 3), or 4) 'Unclear':	1. Risk is low 2. Risk is high 3. Risk unclear
PARTICIPANT SELECTION (1) CONCERNS REGARDING APPLICABILITY	
1) Are the included participants and chosen study setting appropriate to answer the review question, i.e., are the study results generalisable? • This item is not asking whether exclusion of certain participant groups might bias the study's results (as in Risk of Bias above), but is asking whether the chosen study participants and setting are appropriate to answer our review question. Because we are looking to establish test accuracy in both primary presentation and referred participants, a study could be appropriate for 1 setting and not for the other, or it could be unclear as to whether the study can appropriately answer either question • For each study assessed, please consider whether it is more relevant for A) participants with a primary presentation of a skin lesion or B) referred participants, and respond to the questions in either A) or B) accordingly. If the study gives insufficient details, please respond **Unclear** to both parts of the question	**A) For studies that will contribute to the analysis of participants with a primary presentation of a skin lesion (i.e., test naive)** **Yes** – if participants included in the study appear to be generally representative of those who might present in a usual practice setting **No** – if study participants appear to be unrepresentative of usual practice, e.g., in terms of severity of disease, demographic features, presence of differential diagnosis or comorbidity, setting of the study, and previous testing protocols **Unclear –** if insufficient details are provided to determine the generalisability of study participants **B) For studies that will contribute to the analysis of referred participants (i.e., who have already undergone some form of testing)** **Yes** – if study participants appear to be representative of those who might be referred for further investigation. If the study focuses only on those with equivocal lesions, for example, we would suggest that this is not representative of the wider referred population **No** – if study participants appear to be unrepresentative of usual practice, e.g., if a particularly high proportion of participants have been self-referred or referred for cosmetic reasons. Other factors to consider include severity of disease, demographic features, presence of differential diagnosis or comorbidity, setting of the study, and previous testing protocols **Unclear –** if insufficient details are provided to determine the generalisability of study participants
2) Did the study **avoid including** participants with multiple lesions?	**Yes** – if the difference between the number of included lesions and number of included participants is less than 5% **No** – if the difference between the number of included lesions and number of included participants is greater than 5% **Unclear –** if it is not possible to assess
Is there concern that the included participants do not match the review question? 1. If the answer to question 1) or 2) 'Yes': 2. If the answer to question 1) or 2) 'No': 3. If the answer to question 1) or 2) 'Unclear':	1. Concern is low 2. Concern is high 3. Concern is unclear
INDEX TEST (2) RISK OF BIAS (to be completed per test evaluated)	
1) Was the index test or testing strategy result interpreted without knowledge of the results of the reference standard?	**Yes** – if index test described as interpreted without knowledge of reference standard result or, for prospective studies, if index test is always conducted and interpreted prior to the reference standard **No** – if index test described as interpreted in knowledge of reference standard result **Unclear** – if index test blinding is not described
2) Was the diagnostic threshold at which the test was considered positive (i.e., melanoma present) prespecified?	**Yes** – if threshold was prespecified (i.e., prior to analysing study results) **No** – if threshold was not prespecified **Unclear** – if not possible to tell whether or not diagnostic threshold was prespecified
3) For within-person comparisons of index tests or testing strategies (i.e., > 1 index test applied per participant): was each index test result interpreted without knowledge of the results of other index tests or testing strategies?	**Yes** – if all index tests were described as interpreted without knowledge of the results of the others **No** – if the index tests were described as interpreted in the knowledge of the results of the others **Unclear** – if it is not possible to tell whether knowledge of other index tests could have influenced test interpretation **N/A** – if only 1 index test was evaluated
Could the conduct or interpretation of the index test have introduced bias? **For non-comparative and between-person comparison studies** 1. If answers to questions 1) and 2) 'Yes': 2. If answers to either questions 1) or 2) 'No': 3. If answers to either questions 1) or 2) 'Unclear': **For within-person comparative studies** 1. If answers to all questions 1), 2), for any index test and 3) 'Yes': 2. If answers to any 1 of questions 1) or 2) for any index test or 3) 'No': 3. If answers to any 1 of questions 1) or 2) for any index test or 3) 'Unclear':	**For non-comparative and between-person comparison studies** 1. Risk is low 2. Risk is high 3. Risk is unclear **For within-person comparative studies** 1. Risk is low 2. Risk is high 3. Risk is unclear
INDEX TEST (2) CONCERN ABOUT APPLICABILITY	
1) Was the diagnostic threshold to determine presence or absence of disease established in a previously published study? E.g., previously evaluated/established • algorithm/checklist used • lesion characteristics indicative of melanoma used • objective (usually numerical) threshold used	**Yes** – if a previously evaluated/established tool to aid diagnosis of melanoma was used or if the diagnostic threshold used was established in a previously published study **No** – if an unfamiliar/new tool to aid diagnosis of melanoma was used, if no particular algorithm was used, or if the objective threshold reported was chosen based on results in the current study **Unclear** – if insufficient information was reported
2) Were thresholds or criteria for diagnosis reported in sufficient detail to allow replication? Study results can only be reproduced if the diagnostic threshold is described in sufficient detail. This item applies equally to studies using pattern recognition and those using checklists or algorithms to aid test interpretation	**Yes** – If the criteria for diagnosis of melanoma were reported in sufficient detail to allow replication **No** – if the criteria for diagnosis of melanoma were not reported in sufficient detail to allow replication **Unclear** – If some but not sufficient information on criteria for diagnosis to allow replication were provided
3) Was the test interpretation carried out by an experienced examiner?	**Yes** – if the test was interpreted by 1 or more speciality-accredited dermatologists, or by examiners of any clinical background with special interest in dermatology and with

	any formal training in the use of the test **No** – if the test was not interpreted by an experienced examiner (see above) **Unclear** – if the experience of the examiner(s) was not reported in sufficient detail to judge or if examiners were described as 'Expert' with no further detail given **N/A** – if system-based diagnosis, i.e., no observer interpretation
Is there concern that the index test, its conduct, or interpretation differ from the review question? 1. If answers to questions 1), 2), and 3) 'Yes': 2. If answers to questions 1), 2), or 3) 'No': 3. If answers to questions 1), 2), or 3) 'Unclear':	1. Concern is low 2. Concern is high 3. Concern is unclear
REFERENCE STANDARD (3) RISK OF BIAS	
1) Is the reference standard likely to correctly classify the target condition? **A) Disease-positive** – 1 or more of the following: • histological confirmation of melanoma following biopsy or lesion excision • clinical follow-up of benign-appearing lesions for at least 3 months following the application of the index test, leading to a histological diagnosis of melanoma **B) Disease-negative** – 1 or more of the following: • histological confirmation of absence of melanoma following biopsy or lesion excision in at least 80% of disease-negative participants • clinical follow-up of benign-appearing lesions for a minimum of 3 months following the index test in up to 20% of disease-negative participants	**A) Disease-positive** **Yes** – if all participants with a final diagnosis of melanoma underwent 1 of the listed reference standards **No** – If a final diagnosis of melanoma for any participant was reached without histopathology **Unclear** – if the method of final diagnosis was not reported for any participant with a final diagnosis of melanoma or if the length of clinical follow-up used was not clear or if a clinical follow-up reference standard was reported in combination with a participant-based analysis and it was not possible to determine whether the detection of a malignant lesion during follow-up is the same lesion that originally tested negative on the index test **B) Disease-negative** **Yes** – If at least 80% of benign diagnoses were reached by histology and up to 20% were reached by clinical follow-up for a minimum of 3 months following the index test **No** – if more than 20% of benign diagnoses were reached by clinical follow-up for a minimum of 3 months following the index test or if clinical follow-up period was less than 3 months **Unclear** – if the method of final diagnosis was not reported for any participant with benign or non-melanoma diagnosis
2) Were the reference standard results interpreted without knowledge of the results of the index test? Please score this item for all studies even though histopathology interpretation is usually conducted with knowledge of the clinical diagnosis (from visual inspection or dermoscopy or both). We will deal with this by not including the response to this item in the 'Risk of bias' assessment for these tests. For reviews of all other tests, this item will be retained	**Yes** – if the reference standard diagnosis was reached blinded to the index test result **No** – if the reference standard diagnosis was reached with knowledge of the index test result **Unclear** – if blinded reference test interpretation was not clearly reported
Could the reference standard, its conduct, or its interpretation have introduced bias? **For visual inspection/dermoscopy evaluations** 1. If answer to question 1) 'Yes': 2. If answer to question 1) 'No': 3. If answer to question 1) 'Unclear': **For all other tests** 1. If answers to questions 1) and 2) 'Yes': 2. If answers to questions 1) or 2) 'No': 3. If answers to questions 1) or 2) 'Unclear':	**For visual inspection/dermoscopy evaluations** 1. Risk is low 2. Risk is high 3. Risk is unclear **For all other tests** 1. Risk is low 2. Risk is high 3. Risk is unclear
REFERENCE STANDARD (3) CONCERN ABOUT APPLICABILITY	
1) Are index test results presented separately for each component of the target condition (i.e., separate results presented for those with invasive melanoma, melanoma in situ, lentigo maligna, severe dysplasia, BCC, and cSCC)?	**Yes –** if index test results for each component of the target condition can be disaggregated **No –** if index test results for the different components of the target condition cannot be disaggregated **Unclear** – if not clearly reported
2) Expert opinion (with no histological confirmation) was not used as a reference standard 'Expert opinion' means diagnosis based on the standard clinical examination, with no histology or lesion follow-up ***do not complete this item for teledermatology studies	**Yes** – if expert opinion was not used as a reference standard for any participant **No** – if expert opinion was used as a reference standard for any participant **Unclear** – if not clearly reported
3) Was histology interpretation carried out by an experienced histopathologist or by a dermatopathologist?	**Yes –** if histology interpretation was reported to be carried out by an experienced histopathologist or dermatopathologist **No –** if histology interpretation was reported to be carried out by a less experienced histopathologist **Unclear** – if the experience/qualifications of the pathologist were not reported
Is there concern that the target condition as defined by the reference standard does not match the review question? 1. If answers to all questions 1), 2), and 3) 'Yes': 2. If answers to any 1 of questions 1), 2), or 3) 'No': 3. If answers to any 1 of questions 1), 2), or 3) 'Unclear': ***For teledermatology studies only** 1. If answers to all questions 1) and 3) 'Yes': 2. If answers to questions 1) or 3) 'No': 3. If answers to questions 1) or 3) 'Unclear':	1. Concern is low 2. Concern is high 3. Concern is unclear ***For teledermatology studies only** 1. Concern is low 2. Concern is high 3. Concern is unclear
FLOW AND TIMING (4): RISK OF BIAS	
1) Was there an appropriate interval between index test and reference standard? **A)** For histopathological reference standard, was the interval between index test and reference standard ≤ 1 month?	**A)** **Yes** – if study reports ≤ 1 month between index and reference standard **No** – if study reports > 1 month between index and reference standard **Unclear** – if study does not report interval between index and reference standard

B) If the reference standard includes clinical follow-up of borderline/benign-appearing lesions, was there at least 3 months' follow-up following application of index test(s)?	**B)** **Yes** – if study reports ≥ 3 months' follow-up **No** – if study reports < 3 months' follow-up **Unclear** – if study does not report the length of clinical follow-up
2) Did all participants receive the same reference standard?	**Yes** – if all participants underwent the same reference standard **No** – if more than 1 reference standard was used **Unclear** – if not clearly reported
3) Were all participants included in the analysis?	**Yes** – if all participants were included in the analysis **No** – if some participants were excluded from the analysis **Unclear**– if not clearly reported
4) **For within-person comparisons of index tests** Was the interval between application of index tests ≤ 1 month?	**Yes** – if study reports ≤ 1 month between index tests **No** – if study reports > 1 month between index tests **Unclear** – if study does not report the interval between index tests
Could the participant flow have introduced bias? **For non-comparative and between-person comparison studies** 1. If answers to questions 1), 2), and 3) 'Yes': 2. If answers to any 1 of questions 1), 2), or 3) 'No': 3. If answers to any 1 of questions 1), 2), or 3) 'Unclear': **For within-person comparative studies** 1. If answers to all questions 1), 2), 3), and 4) 'Yes': 2. If answers to any 1 of questions 1), 2), 3), or 4) 'No': 3. If answers to any 1 of questions 1), 2), 3), or 4) 'Unclear':	**For non-comparative and between-person comparison studies** 1. Risk is low 2. Risk is high 3. Risk is unclear **For within-person comparative studies** 1. Risk is low 2. Risk is high 3. Risk is unclear
BCC = basal cell carcinoma; cSCC = cutaneous squamous cell carcinoma.	

8 Summary study details – in-person evaluations for detection of invasive melanoma and intraepidermal melanocytic variants

Study author **Pathway (clear/unclear)** **Outcomes**	**Study type** **Country** **Setting** **Pts / lesions**	**Inclusion criteria**	**Index tests (algorithm)** **Diagnostic approach**	**Threshold**	**Observer qual. (n)** **Experience**	**Reference standard** **Final diagnoses** **Prevalence (MM+MiS)**	**Exclusions**
Ahnlide 2016 5-Referred (selected on reference) (u) MM+MiS	WPC-algs R-CS Secondary Sweden NR / 309	Excised melanocytic skin lesions with recorded dermoscopy ABCD score and clinician's preliminary diagnosis. Preliminary diagnosis of LM or SN excluded.	Dermoscopy 1. no algorithm 2. ABCD In-person	1. Subjective impression (dx of MM) 2. >4.75; >5.45	Dermatology registrar or consultants (n=13; experienced unit; dermoscopy training); visiting residents data excluded	Histology MM 23; MiS 23 BN: 263	57 lesions with missing scoring; 5 non-melanocytic diagnosis; 5 with pre-op dx of LM or SN; 1 with ambigious histopathological diagnosis
Bauer 2000 4-Referred (c) MM+MiS	WPC NR-CS Secondary Italy 311 / 315	PSL examined during a campaign for the early diagnosis of cutaneous melanoma (CM)	Dermoscopy (no algorithm) [Also evaluated CAD-Dermoscopy)	NR Subjective impression (dx of malignancy)	Dermatologist (n=3; trained in recognition of PSLs) Consensus of 3 (expert consult for disagreements)	Histology MM 30; MiS 12 'Atypical' dysplastic 25; BN 212; NML 36	
Benelli 1999 5-Referred (selected on reference) (u) MM+MiS	WPC P-CS Italy Secondary NR / 401	All PSL observed and excised at the Dermatologic Surgery Department	1. VI (ABCDE) 2. Dermoscopy (7FFM) In-person	1. >/=1 to all 5 chars present 2. Score >=2	Dermatologist (n=2; exp NR) Consensus of 2	Histology MM 54; MiS 6 BCC 1 BN 337; LS 5; SK 1 60/401; 15%	None reported
Bono 2002 5-Referred (selected on reference) (c) MM+MiS	WPC P-CS Italy Specialist clinic 298 / 313	PSL with a more or less important suspicion for MM on VI and/or dermoscopy	1 .VI (no algorithm) 2. Dermoscopy (no algorithm) In-person [also evaluates CAD-Dermoscopy]	VI - subjective impression Dermosc - >=1 char present	Surgical oncologist (n=4; high) Single obs	Histology MM 55; MiS 11 BCC 6; 8 SK; 3 SN; BN 230 66/313; 21%	None reported
Bono 2002b 5-Referred (selected on reference) (c) MM+MiS	WPC P-CS Italy Specialist clinic 157 / 161	PSL <=6mm requiring surgical biopsy for diagnosis based on clinical or dermoscopic suspicion of MM	1. VI (no algorithm) 2. Dermoscopy (no algorithm) In-person	VI - subjective impression Dermosc - >=1 char present	Surgical oncologist (n=2; high) Single obs	Histology MM 10; MiS 3 BCC 2; SK 4; SN 5; BN 124 13/161; 8%	None reported
Bono 2006 5-Referred (selected on reference) (c) MM+MiS	WPC R-CS Italy Specialist clinic	PSL <=3mm undergoing excision due to a more or less important suspicion for MM on VI and/or dermoscopy	1. VI (no algorithm) 2. Dermoscopy (Menzies) In-person	VI - subjective impression Dermosc - NR	NR; assumed surgical oncologist as per Bono 2002; Bono 2002b (n=4; exp NR) Single obs	Histology MM 19; MiS 4 SN 3; BN 169; Other 11 23/206; 11%	None reported

	204 / 206						
Broganelli 2005 5-Referred (selected on reference) (u)	NC P-CS Secondary Italy NR / 638	PSL undergoing excision; 2x2 for melanocytic only included	Dermoscopy (7PCL) Unclear if in-person or image-based	> 1 change in minor criteria or >= 1 major char present	Dermatologist (assumed) (n=NR; exp NR)	Histology MM+MiS 108 'Non-melanoma' 530	
Carli 1994 5*-Equivocal (selected on reference)(c) MM+MiS	NC NR-CS Secondary Italy 67 / 67	Clinically suspicious melanocytic lesions undergoing excision for diagnostic purposes (obvious MM excluded)	Dermoscopy (pattern analysis) In person	Irregular pigmented network plus >=1 other listed characteristic	Dermatologist (n=2; exp High) Consensus of 2	Histology MM 3; MiS 2 BN 62 5/68; 7%	
Carli 2002 5-Referred (selected on reference) (u) MM+MiS	WPC R-CS Italy Secondary NR / 256	Clinically equivocal and suspicious PSL subjected to excisional biopsy at the Institute of Dermatology	1. VI (no algorithm) 2. Dermoscopy (pattern) In-person (Dermoscopy – image-based)	Subjective impression	Dermatologist (n=2; High exp – "extensive experience in both clinical and dermoscopic diagnosis") Consensus of 2	Histology MM 40; MiS 14 BCC 5 BN 177; SN 16; SK 4 54/256; 21%	None reported
Cristofolini 1994 5-Referred (selected on reference) (u) MM+MiS	WPC P-CS Italy Secondary NR / 220	Patients with PSL presenting during a campaign for the early diagnosis of cutaneous melanoma at the Dermatology Department	1. VI (ABCDE) 2. Dermoscopy (pattern) In-person	1. ≥2 chars present 2. >=1 char present	Dermatologist (n=4; High exp - dermatologists had all been trained in the recognition of pigmented lesions) Unclear obs interp	Histology MM+MiS 33 BCC 0 BN 181; SK 4; 2 other 33/220; 15%	None reported
Dreiseitl 2009 4-Referred (c) MM+MiS	NC P-CS Specialist clinic Austria 458 / 3021	Patients presenting at PSL clinic which serves as a secondary and tertiary referral centre	Dermoscopy (no algorithm)	NR Subjective impression (dx of MM)	Dermatologist (n=1; 'expert') Single observer	Histology or FU (6 mos) MM+MiS 31 (27 pts) 'Benign': 2990 (431 pts) 27/458; 6%	806 lesions (53 patients) with inadequate follow-up
Duff 2001 5-Referred (selected on reference) (c)	NC R-CS Specialist clinic UK NR / 2372	Excised lesions recorded on PSL database	Dermoscopy (no algorithm) In person	Subjective assessment (decision to excise)	Plastic surgeon (n=1; exp NR) Single observer	Histology or FU MM 400; MiS 186; BCC: 316; cSCC: 97 Dysplastic 195; "other" 14; 'Benign' (not excised): 1164 586/2372; 25%	None reported Results for BCC; SCC not disaggregated from benign lesions
Durdu 2011 5-Referred (selected on reference) (u) MM+MiS Any	WPC P-CS Secondary Turkey 176 / 200	PSL that could not be diagnosed with only dermatologic physical examination; 2x2 included for melanocytic subset	Dermoscopy (ABCD; nonmelanocytic excluded first) [Also evaluated exfoliative cytology] In person	NR	Dermatologist (n=1; exp NR) Single observer	Histology MM+MiS 10; BCC: 34; Other malignant 2 SK 24; BN 100; DF 12; Warts 16; Dirt 1; Other 1 10/200; 5%	-
Feldmann 1998 5-Referred (selected on reference) (u) MM MM+MiS	NC P-CS Secondary Austria NR / 500	Melanocytic lesions examined by dermatoscopy prior to excision	Dermoscopy (ABCD) In person	>5.45	NR(n=NR; exp NR) Unclear obs interp	Histology MM 25; MiS 5 BN 272; dysplasia 190; lentigines 7; lentigo nevi 1 30/500; 6%	NR
Gokdemir 2011 5-Referred (selected on reference) (u) MM+MiS	NC NR-CS Secondary Turkey 362 / 449	Patients with melanocytic and non-melanocytic skin lesions with dermoscopic and histologic diagnoses.	Dermoscopy (no algorithm) Unclear if in-person or image-based	Subjective assessment (dx of MM)	Dermatologist (n=NR; exp High "at least 2 years' experience with Molemax II") Unclear obs interp	Histology MM+MiS 13; BCC: 45 Benign: 390 13/448; 3%	Bham team: 1 BCC moved from FP to TN]
Grimaldi 2009 2-Limited prior testing (c) MM+MiS	WPC P-CS Italy Primary 197/235	Cutaneous PSL requiring confirmation of diagnosis by teledermatology.	1. VI (no algorithm) 2. Dermoscopy (no algorithm) In-person (Single) [Also evaluated Teledermatology]	Subjective impression ('suspicious for malignancy')	GP (n=13) Assumed Low (Expertise NR; simple protocols for diagnosis provided for study purposes)	Histology/Clinical FU (6mos) MM+MiS 5; BCC 0; Benign 230 (NR) 5/235; 2%	NR
	WPC	Lesions excised on the basis of clinical suspicion (history,	Dermoscopy (pattern	Subjective assessment (dx	Dermatologist (n=2; exp High)	Histology	Only 50% of imaged nevi were included (randomly

Guitera 2009a (Modena) 5-Referred (selected on reference) (c) MM+MiS	P-CS Secondary Italy 195 / 195	dermoscopy examination, and/or digital monitoring)	analysis) [Also evaluated RCM] In-person	of MM)	Single observer	MM 61; MiS 18 BN 116 (incl 22 SN) 79/198; 41%	selected from the image database prior to analysis) to reduce the MM/nevus ratio
Haenssle 2010a (FV) 4-Referred (u) MM+MiS	NC P-CS Secondary Germany 688 / 11137 FV: 8449	Participants at increased risk for melanoma: >50 common and/or <=3 atypical nevi; atypical mole syndrome (AMS); or familial atypical mole and multiple melanoma syndrome. [first visit data included here]	Dermoscopy (7PCL)	>=3	Dermatology residents (n=13; formally trained in dermoscopy and supervised by experienced dermatologist) Consensus of 2	Histology or FU (every 3, 6, or 12 mos) Full sample: MM 77; MiS 50; BCC 2 BN 1047; SN 16; SK 12; Other benign 9935 40/8449; 0.005%	
Haenssle 2010b (FU) 7–Follow-up (u) MM+MiS	NC P-CS Secondary Germany Full sample; 688 / 11137 FU: 2688 lesions	Participants at increased risk for melanoma: >50 common and/or <=3 atypical nevi; atypical mole syndrome (AMS); or familial atypical mole and multiple melanoma syndrome [FU data only included here]	Dermoscopy (7PCL) In person	>=3	Dermatology residents (n=13; formally trained in dermoscopy and supervised by experienced dermatologist) Consensus of 2	Histology or FU (every 3, 6, or 12 mos) Full sample: MM 77; MiS 50; BCC 2 BN 1047; SN 16; SK 12; Other benign 9935 87/2688; 3%	
Kittler 1999 5-Referred (selected on reference) (u) MM+MiS	WPC-algs P-CS Secondary Austria 352 / 373	Melanocytic PSL < 1cm in diameter, consecutively excised	Dermoscopy 1. ABCD 2. ABCDE (developed in this study) In-person	1. Sensitivity at range of specificities (randomly sampled 75% spec) [author comm. Suggests >4.75 used] 2. cutoffs between 1.30 to 7.35	Dermatologist (assumed) (n=NR; exp NR) Unclear obs interp	Histology MM 55; MiS 18 SK 4; BN 126; atypical nevi 113; congenital nevi 3, SN 13; blue nevi 7; solar lentigines 14; DF 1; combined nevi 2 73/356; 21%	Non melanocytic lesions (n=17; including angiomatous tumours, pigmented SK, DF, and pigmented BCC) 'easily distinguished by standard ELM criteria and pattern analysis'
Langley 2007 5-Referred (selected on reference) (u) MM+MiS	WPC P-CS Specialist clinic Canada 125 / 125	Patients with lesions scheduled for excision at the PLC to either remove atypical nevi or to rule out melanoma or for cosmetic reasons; excluded if lesion not amenable to RCM or prior dx biopsy	Dermoscopy (pattern analysis) [Also evaluated RCM] In-person	Subjective assessment (dx of MM)	Dermatologist (assumed); (n=1; exp NR)	Histology MM 22; MiS 15 BN 88 37/125; 30%	Technical difficulties with imaging (n=2)
Menzies 2009 2-Limited prior testing (c) MM+MiS Any	WPC-algs P-CS Aus. Primary NR/374	PSL that would be biopsied or referred on after routine naked eye examination	1 .VI (no algorithm) 2. Dermoscopy (no algorithm) In-person (Single)	Subjective impression ('correct diagnosis of melanoma'; excise decision)	GP (n=62) Assumed Low (trained in dermoscopy for study; required history of excision or referral of at least 10 pigmented skin lesions over the previous 12-month period but no prior dermoscopy use)	Histology/Clinical FU (3-6 mos)/Expert dx MM+MiS 32; BD 2; Benign 323; Unknown 9 4%	6 BCC and 2 BD excluded by authors, 43 excluded as both VI +Dermoscopic diagnoses not available
Morales Callaghan 2008 5-Referred (selected on reference) (u) MM+MiS	WPC P-CS Spain Secondary 166 / 200	Randomly selected melanocytic lesions; melanocytic on both clinical and dermoscopic criteria	1. VI (no algorithm) 2. Dermoscopy (no algorithm) In-person	NR	Dermatologist (n=2; high exp – "experience in dermoscopy") Consensus of 2	Histology MM+MiS 6 BN 184; SN 1; Other 9 6/200; 3%	None reported
Nachbar 1994 5-Referred (selected on reference) (u) MM+MiS	NC P-CS Secondary Germany NR / 194	Pigmented melanocytic skin lesions consecutively excised	Dermoscopy (ABCD) In person [excluded VI data as dermoscopy also used for VI]	>5.45	Dermatologist (assumed) (n=NR; exp High)	Histology MM+MiS 69; BCC 3 BN 103; SK 19 69 / 194; 36%	
Soyer 1995 5*-Equivocal (selected on reference)(c) MM+MiS Any	WPC NR-CS Austria NR / 159	PSL difficult to diagnose on clinical grounds alone	1. VI (no algorithm) 2. Dermoscopy (pattern) In-person	NR	Dermatologist (n=2; exp High; "the examination was performed by a dermatologist expert in dermoscopy") Single obs	Histology MM 50; MiS 15 BCC 3; SK 18; AK 4; BN 61; Other 7 65/159; 41%	None reported
Stanganelli 2000 4-Referred (c) MM+MiS Any	WPC R-CS Italy Specialist clinic	PSL referred by dermatologists and general practitioners either for pre-surgical assessment or consultation	1. VI (ABCD) 2. Dermoscopy (no algorithm) In-person (Single)	NR Subjective impression	NR (assumed dermatologist - described as one of the co-authors; n=1)	Histology / Registry FU MM+MiS 55 BCC 43; Benign 3274 55/3372; 2%	None reported BCC: 3 BCCs considered to be MM were classed as TN rather than FP for review purposes

	NR / 3372						

Footnotes:

1 Test naïve; 2 Limited prior testing; 3 Limited prior testing (with selection on reference standard); 3* Limited prior testing (with selection on reference standard and equivocal nature of lesions); 4 Referred for further assessment; 5 Referred for further assessment (with selection on reference standard); 5* Referred for further assessment (with selection on reference standard and equivocal nature); 6 Referred for further assessment (equivocal on specialist review); 7 Lesions that have been undergoing follow-up

c- clearly positioned on clinical pathway; u – unclear position on clinical pathway; NR – not reported; PSL – pigmented skin lesion; PLC – pigmented lesion clinic; MM – malignant melanoma; MiS – melanoma *in situ* (or lentigo maligna); BCC – basal cell carcinoma; cSCC – cutaneous squamous cell carcinoma; LS – lentigo simplex; SK – seborrheic keratosis; SN – Spitz nevi; AK – actinic keratosis; BN – benign naevi; BD – Bowen's disease; DF – dermatofibroma; FU – follow-up; R –retropsective; P – prospective; CS – case series; CCS – case control study; WPC – within person comparison (of tests); BPC – between person comparison (of tests); NC – non comparative; RCM – reflectance confocal microscopy; CAD – computer-assisted diagnosis

9 Summary study details – image-based evaluations for detection of invasive melanoma and intraepidermal melanocytic variants

Study author **Pathway (clear/unclear)** **Outcomes**	**Study type** **Country** **Setting** **Pts / lesions**	**Inclusion criteria**	**Index tests (algorithm)** **Diagnostic approach**	**Threshold**	**Observer qual. (n)** **Experience**	**Reference standard** **Final diagnoses** **Prevalence (MM+MiS)**	**Exclusions** **Comments**
Alarcon 2014 5*-Equivocal (selected on reference) (c) MM+MiS	WPC P-CS Specialist clinic Spain 264 / 264	Dermoscopically equivocal pigmented lesions, assumed to be melanocytic	Dermoscopy (No algorithm) Image-based (RCM, site, age) [also evaluated RCM]	NR; dx of MM	Dermatologist (n=3; exp NR) described as expert in RCM Single observer	Histology or FU ;79 followed-up MEL 92; BCC: 12 BN 107; 53 SK and AK 92/343; 27%	79 lesions without criteria of malignancy on RCM were scheduled for clinical or digital follow-up.
Annessi 2007 5*-Equivocal (selected on reference) (u) MM+MiS	WPC-algs NR-CS Specialist clinic Italy 195 / 198	Atypical macular melanocytic lesions; all >5 mm diameter, with a flat or barely elevated surface and at least 3 of the following features: (a) asymmetry, (b) irregular margins, (c) ill-defined borders, and (d) color variegation.	Dermoscopy (Pattern analysis; 7PCL; ABCD) Image-based (blinding NR)	NR - likely 'standard'; ABCD >=4.75	Dermatologist (n=2; exp High) ELM-experienced dermatologists" Consensus of 2	Histology MM 72 ; MiS 24 BN 102 96/198; 48%	None reported
Argenziano 1998 5-Referred (selected on reference) (u) MM+MiS	WPC-algs; obs R-NR Secondary Italy NR / 342	Atypical melanocytic skin lesions with dermoscopic images that had undergone biopsy due to clinician suspicion	Dermoscopy (pattern analysis; ABCD; 7PCL) Image-based (blinded)	Overall dx MM; ABCD >4.75; 7PCL >=3	Dermatologist (n=3 experienced; n=2 less experienced who underwent training) Consensus of 2 (expert) Single (less experienced)	Histology MM 99; MiS 18 BN 225 117 / 342; 34%	None reported
Argenziano 2011 5-Referred (selected on reference) (c) MM+MiS	WPC-algs CCS Secondary Italy NR / 300	Randomly sampled 100 melanomas; 100 excised BN 100 BN that showed no relevant changes to warrant excision during the FU period; all <=15mm	Dermoscopy (pattern analysis; 7PCL; 7PCLrev) Image-based (blinded)	Pattern – dx of MM and excise decision; 7PCL >=3; rev >=1)	Dermatologist (n=8; exp NR) average	Histology or FU MEL 100 57 Clark nevi, 28 SN, 10 congenital naevi, 5 blue naevi; 100 not excised 100/300; 33%	NR
Benelli 2000 5-Referred (selected on reference) (u) MM+MiS	WPC CCS Secondary Italy NR / 600	All small (<= 6 mm) melanomas and melanocytic nevi consecutively excised over two different time periods	1. VI (ABCD) 2. Dermoscopy (7FFM) Image-based (blinding NR)	Both >=2	Dermatologist (assumed) (n=3; exp NR) evaluated by 3 different observers; in case of disagreement , the majority view prevailed Consensus of 3	Histology alone MEL 76 BN 524 76 / 600; 13%	NR
Benelli 2001 5-Referred (selected on reference) (u) MM+MiS	WPC R-CS Italy Training images NR / 50	Slides of PSL selected for evaluation during a training course on dermoscopy. Lesions not located on head, palms or soles	1. VI (ABCDE) 2. Dermoscopy (7FFM) Image-based (blinded)	1. >=3 & >=2 2. >=2	Expert author (n=1); Dermatologists (n=65) Single author - High exp; Average result for dermatologist group; exp NR	Histology MM 10, MiS 2 BCC 2 BN 25, SN 5, SK 3, Other 2 (1 missing) 12/50; 24%	None reported
Binder 1994 5-Referred (selected on reference) (u) MM+MiS	WPC RCS Secondary Austria NR / 200	Images of PSL randomly selected from a image database.	Dermoscopy (pattern analysis) [Also evaluates CAD dermoscopy] Image-based (blinded)	Subjective impression (dx of MM)	Dermatologist (n=3; exp High) Consensus of 2	Histology MEL 40 BN 60 40/100; 40%	None reported
Binder 1995 5-Referred (selected on reference) (u) MM+MiS	WPC-obs RCS Secondary Austria NR / 240	PSL with available dermoscopy images, both with and without oil immersion, and histological confirmation of diagnosis.	Dermoscopy (no algorithm) [Dermoscopy with/without oil immersion] Image-based (blinded)	Subjective impression (dx of MM)	Dermatologist (n=6 expert; n=13 non-expert); Average	Histology MEL 57; BCC: 8 Severe dysplasia: 42; other 'Benign' : 133	None reported Test results not disaggregated for BCC

						57/240; 24%	
Binder 1999 5-Referred (selected on reference) (u) MM+MiS	WPC-algs RCS Secondary Austria NR / 250	Randomly selected, histologically proven PSL with digital dermoscopy images	Dermoscopy (pattern analysis; ABCD) Image-based (blinded)	Subjective impression (dx of MM); ABCD at >4.75; >5.45	Mixed (n=17; exp mixed) Dermatology residents - 5; Dermatologist (board-certified) - 12 Average result	Histology MM 34; MiS 7 BN 182; 13 SN; 14 lentigines 41/250; 16%	None reported
Blum 2003 5-Referred (selected on reference) (u) MM+MiS	WPC-alg R-CS Specialist clinic Germany NR / 269	Melanocytic skin lesions to be excised because of clinically and/or dermoscopically clear or suspicious malignancy, or by the wish of the patient after clear benign diagnosis*	ABCD (modified); ABCDE (modified) Image-based (unclear)	NR	Dermatologist (assumed) (n=NR; exp NR) NR	Histology MM 71; MiS 9; LM 4 'Benign': 185 84/269; 31%	*dataset overlaps Blum 2004b so not included in primary analysis, only algorithm comparisons (recruited November 1998 to March 2000)
Blum 2003b 5-Referred (selected on reference) (u) MM+MiS	NC R-CS Specialist clinic Germany 205/254	All lesions of patients with multiple atypical naevi excised due to suspicious clinical and/or dermoscopic features were included*	New (based on Hofmann-Wellenhof 2001) Image-based (blinded)	Presence of reticular, globular and homogeneous structures	Dermatologist (assumed) (n=2; exp NR) Consensus of 2	Histology MM 63; MiS 12 BN 64; Dysplastic 96; other nevus 19 75/254; 30%	*dataset overlaps Blum 2004b so not included in primary analysis, only algorithm comparisons (recruited September 1998 to December 1999)
Blum 2004 5-Referred (selected on reference) (u) MM+MiS	WPC-obs R-CS Specialist clinic Germany 157/157	PSL excised due to suspicious clinical and / or dermoscopic features	Pattern analysis Image-based (blinded)	Level of suspicion 'roughly 50% or more'.	Dermatologist (assumed) (n=3; with “different experiences in dermoscopy: excellent (A), average (B) and beginner (C)." Single	Histology MM 29;MiS 2 BN 53; dysplastic 59 'epithelial benign’ 13 32/157; 20%	*dataset overlaps Blum 2004b so not included in primary analysis, only observer comparisons (recruited September 1998 to March 1999)
Blum 2004b 4-Referred (u) MM+MiS	WPC-algs P-CS Specialist clinic Germany NR / 837	Melanocytic skin lesions imaged prospectively at the Pigmented Lesion Clinic	Dermoscopy (ABCD; 7PCL; 7FFM; Menzies) Image-based (blinded) [also evaluated CAD-Dermoscopy]	NR - author confirms 'published standards used'	Dermatologist (assumed); n=1 Single observer	Histology or FU (568 benign examined 2-3 times in 6 months) MM 71; MiS 9; LM 4 'Benign' 766 84/837; 10%	None reported
Bourne 2012 3–Limited testing (selected on reference) (c) MM+MiS	WPC-algs; algs R-CS Aus. Primary 46 / 50	All skin lesions excised to exclude skin cancer (and 3 examples common lesions assessed as clearly benign and not biopsied)	VI (no algorithm) Dermoscopy (3-point; Menzies; BLINCK*) Image-based (blinded)	NR	GP (n=3) Clinical nurse (n=1) Mixed exp “varying levels of dermatoscopic experience”. Average	Histology / Clin FU / Expert dx MM 1; MiS 8 BCC 6; SK 5; BN 11; Other 19 9/45; 20%	5 non-pigmented specimens (not further identified) in the set of 50 were excluded from dermoscopic evaluations *data for BLINCK excluded as derivation
Carli 2002 5-Referred (selected on reference) (u) MM+MiS	WPC NR-CS Secondary Italy NR / 256	Clinically equivocal and suspicious PSL	1. VI (no algorithm) (in-person) 2. Dermoscopy (pattern analysis) (in-person and image-based) Image-based (age, site provided)	Subjective impression (dx of MM)	Dermatologist (n=2; exp High; ‘extensive experience in both clinical and dermoscopic diagnosis of PSLs’) Consensus of 2	Histology alone MM 40; MiS 14 BCC: 5 SK 4; BN 168; 9 blue naevi; 16 SN 54/256; 21%	None
Carli 2002b 5-Referred (selected on reference) (u) MM+MiS Any	WPC R-CS Italy Secondary NR / 57	Clinically suspicious or equivocal PSL undergoing excision for diagnostic purposes; all <= 14mm diameter	1. VI (NR) 2. Dermoscopy (NR) Image-based (blinded)	NR	Dermatologists (n=2) High exp ('with experience in the field of '); consensus of 2	Histology MM 6, MiS 5 BCC 10 BN 31, SK 1; Other 4 11/57; 19%	4 ‘not evaluables’ excluded (NB these differ between clinical images and dermoscopic images (1 MM excluded from VI analysis)
Carli 2003 5-Referred (selected on reference) (u) MM+MiS	WPC-algs RCS Secondary Italy NR / 200	Melanocytic lesions <14 mm in diameter, excised because they were clinically suspicious or equivocal	Dermoscopy (pattern analysis; ABCD; 7PCL) Image-based (blinded)	Subjective impression (dx of MM); ABCD >5.45; 7PCL>=3	Dermatology registrar (n=5; exp low) Single observer	Histology MM 30; MiS 14 BN 156 44/200; 22%	None reported
Carli 2003b 5*-Equivocal (selected on reference) (u) MM+MiS	WPC R-CS Italy Secondary NR / 200	Clinically difficult to diagnose or equivocal melanocytic lesions randomly selected; all melanomas <1mm thickness.	1. VI (no algorithm) 2. Dermoscopy (own choice) Image-based (blinding NR)	subjective impression	Dermatology registrar (n=2); dermatologists (senior experts n=2; practicing dermatologists n=4) Average result	Histology MM 40; MiS 24 BN 136 64/200; 32%	None reported
Carrera 2016 4-Referred (u) MM+MiS	WPC-algs CCS Specialist clinic	Images of melanocytic lesions including MM with unequivocal histology, and histologically verified nevi or nevi demonstrating stability under sequential dermoscopic imaging over time.	Dermoscopy (ABCD; 7PCL; CASH; Menzies; 3PCL; Chaos/clues)	>4.75; >=3; >=6; 2neg1pos; >=2; both present	GP 24; Derm registrar 25; Dermatologist 73; 1 medical student and 7 'other';	Histology or FU (Sequential dermoscopic imaging over time; n=NR)	None reported * Up to 50 lesions per PLC (1:3 ratio of MEL to BN; 1:1 polarized or nonpolarized

	Multicentre NR / 477*		Image-based (clinical image also provided)		Mean 12y (SD 8.7) dermatology exp; 93.8% 'comfortable' using dermoscopy Consensus (>=50%)	MEL 119 BN: 358 119/477; 25%	images); randomized into 12 image sets of 39 (n = 8) or 40 (n = 7) unique lesions and 5 nonunique lesion images (2 MEL, 3 BN) repeated in all sets.
Dal Pozzo 1999 5-Referred (selected on reference) (u) MM+MiS	NC PCS Secondary Italy NR / 713	PSL observed clinically and dermoscopically	Dermoscopy (7FFM) Image-based (blinded)	>=2	Dermatologist (assumed) (n=3; exp NR) Consensus of 3	Histology MM 139; MiS 29; BCC: 1 SK 3; BN 536; Other 5 168/713; 24%	None All BCC considered TN
Di Meo 2016 5-Referred (selected on reference) (u) MM+MiS	WPC-algs RCS Secondary Italy 125 / 125	Melanocytic skin lesions that underwent excision (*accuracy data excludes the dysplastic nevi)	Dermoscopy (3PCL; CASH; 4PCL) Image-based (blinded)	>=2 chars present; >=8; >=2	Dermatologist (n=2; exp High) NR	Histology ; MEL 32 BN 43 32/75; 43%	50 lesions with mild/moderate dysplasia excluded
Dolianitis 2005 5-Referred (selected on reference) (u) MM+MiS	WPC-algs R-CS Multi-centre Training images NR / 40	Melanocytic skin lesions randomly selected from a collection of dermoscopic images belonging to one author.	1. VI (no algorithm) 2. Dermoscopy (Pattern analysis; Menzies Criteria; 7 point; ABCD) Image-based (blinded)	1. subjective impression 2. subjective impression; NR; NR; >4.75	Dermatologists(n=16); dermatology trainees (n=16); GPs (n=35). Mixed exp ("range of experience levels with assessment of skin lesions"); Average result	Histology (n=39); Expert diagnosis (n=1) MM 18, MiS 2 BN 12; SN 3; Other 4 20/20; 50%	None reported; poor quality images exclusion criterion
Dummer 1993 5*-Equivocal (selected on reference) (c) MM+MiS	WPC P-CS Secondary Germany NR / 771	Patients with skin lesions difficult to diagnose clinically	1. VI (no algorithm) 2. Dermoscopy (Pattern analysis) Image-based (blinding NR)	Unclear (German language); dx of MM	Dermatologist (assumed) (n=2; exp Unclear) limited detail; German paper Single	Histology MM 23 BN 706; SK 4; BMN 32 23/771; 3%	Further 53 non-melanocytic lesions not included prior to examination (no melanomas present in this group)
Feci 2015 5-Referred (selected on reference) (u) MM+MiS	BPC RCS Secondary Italy 321 / 321	PSL suspicious for melanoma and excised; observers randomly allocated to observation with different 'stressors'*	Dermoscopy (pattern analysis) Image-based (blinded)	NR; dx of MM	Dermatologist (n=3; exp High) 'expert dermatologists' 'with at least 10 years' exp in dermoscopy' NR	Histology MM 99; MiS 33 BN 219 34/107; 32%	None reported * Data pooled across arms for primary analysis
Feldmann 1998 5-Referred (selected on reference) (u) MM MM+MiS	NC P-CS Secondary Austria NR / 500	Melanocytic lesions examined by dermatoscopy prior to excision	Dermoscopy (ABCD) In person	>5.45; >4.2	NR(n=NR; exp NR) Unclear obs interp	Histology MM 25; MiS 5 BN 272; mild/ moderate dysplasia 190; lentigines 7; lentigo nevi 1 30/500; 6%	NR
Ferrari 2015 5*-Equivocal (selected on reference) (u) MM+MiS	WPC R-CS Secondary Italy NR / 322	Melanocytic lesions with equivocal clinical and/or dermoscopic features that underwent excision	Dermoscopy (7-point) Image-based (RCM, image) [also evaluated RCM in subgroup]	>=3; dx of MM	Dermatologist (n=1; exp NR) Single	Histology MEL 70 'Benign' nevi: 252 (including 15 SN) 70/322; 22%	90 'positive-clear cut' lesions scoring 5 or more were excluded from RCM evaluation
Ferris 2015 5-Referred (selected on reference) (u) MM+MiS	WPC-obs R-NR Secondary US NR / 65	Dermoscopic images of skin lesions excised on the basis of clinical suspicion of malignancy, with available histologic diagnoses	Dermoscopy (no algorithm) Image-based (blinded) [Also evaluates CAD-Dermoscopy]	NR; excise decision	Dermatologist (n=2 board certified); Derm residents (n=10); Physician assistants practicing in dermatology (n=8) Average per group	Histology MM 15; MiS 10 BN 20, blue nevi 2, lentigines 4 , SK 4 25/65; 38%	None reported
Friedman 2008 5-Referred (selected on reference) (u) MM MM+MiS	WPC CCS Secondary/Private US 94 / 99	An industry database of images of PSL <=6mm was used to sample images of melanoma and non melanoma lesions; high-grade dysplastic nevi were excluded.	Dermoscopy (no algorithm) Image-based (site, age, gender) [Also evaluates CAD-Spectroscopy]	Correct dx;; excise decision	Mixed - sec (n=10; exp High) Average result (reports mean and median; mean used)	Histology MM 21; MiS 28; BCC: BN 34; SK 2; 14 other benign 49/99; 49%	None reported
Gereli 2010 5-Referred (selected on reference) (u) MM+MiS	WPC-algs CCS Secondary Turkey NR / 96	Lesions considered clinically atypical* before dermoscopic examination and excisional biopsy.	Dermoscopy (3PCL; 7PCL) Image-based (blinded)	>=2 chars present; >=3	Dermatologist (assumed) (n=3; exp mixed) "two experienced and one inexperienced observers" Average result	Histology MM 44 MiS 4 SK 2; Blue nevi 2; BN 44 48/96; 50%	None reported (*determined by >=3 of: diameter >5 mm, ill-defined borders, irregular margins, presence of papular and macular components).
Gilmore 2010	NC R-CS	Polarised dermoscopic images of atypical melanocytic lesions	Dermoscopy (no algorithm)	NR; excise decision	Dermatologist (assumed) (n=1; exp	Histology MEL 36	130 in derivation set of lesions

5-Referred (selected on reference) (u) MM+MiS	Secondary Austria NR / 69		Image-based (blinded)		High) Single observer	BN (dysplastic): 33 36/69; 52%	
Glud 2009 5-Referred (selected on reference) (u) MM+MiS	WPC P-CS Secondary Denmark 65 / 83	Patients referred for excision biopsy of where the diagnosis of melanoma could not be excluded on clinical investigation	Dermoscopy (no algorithm) Image-based (blinded) [Also evaluates CAD spectroscopy]	NR; dx of MM	Dermatologist (n=1; exp High) Single observer	Histology MM 7; MiS 5; 1 mel mets (incl as D-) SK 1; BN 57; BD 1; DF 6; Other 5 12/83; 14%	None reported
Guitera 2009b (Sydney) 5-Referred (selected on reference) (u) MM+MiS	WPC P-CS Specialist clinic Australia 131 / 131	Lesions excised on the basis of clinical suspicion (history, dermoscopy examination, and/or digital monitoring)	Dermoscopy (pattern analysis) Image-based (age, site) [Also evaluates RCM]	NR; dx of MM	Dermatologist (n=2; exp High; 'expert') Single observer	Histology MM 28; MiS 16 BN 87 (incl 3 SN) 44/131; 34%	(25 lesions out of 156 were rejected for poor quality dermoscopy image, blinded to the diagnostician)
Hauschild 2014 5-Referred (selected on reference) (u) MM MM+MiS	WPC-obs CCS* Secondary/Private US 130 / 130	Subset of PSL evaluated in a MelaFind study (Monheit 2011); 65 melanoma and 65 non-melanoma randomly selected. Excluded ulcerated, non-pigmented, or located on excluded anatomic sites.	Dermoscopy (no algorithm) [Also evaluates CAD spectroscopy] Image-based (clinical image, pt history)	NR; excise decision	Dermatologists (n=202; randomised between 2 arms); PSL experts (n=9) Single observer	Histology MM 36; MiS 29 'Benign' diagnoses: 65 65/130; 50%	*RCT of dx based on clinical/dermoscopic images versus same plus MelaFind, with observers randomised between arms
Kittler 1998 5*-Equivocal (selected on reference) (u) MM+MiS	NC NR-CS Secondary Austria NR / 50	PSL images selected on image quality and difficulty of diagnosis; all melanomas has "only subtle ELM features as clues to the malignancy of the lesion .. difficult to differentiate from benign"	Dermoscopy (No algorithm); compared photographic slides and compressed digital images; latter used for review Image-based (blinded	Subjective impression; dx of MM	Dermatologist (n=8; exp NR) described as 'pre-trained in ELM' Single (randomly sampled one for inclusion)	Histology MEL: 23 SK 1; BN 26 23/50; 46%	None reported
Kittler 2001 7-Follow-up (u) MM+MiS	NC CCS Secondary Austria 20 / 80	Images retrieved from a PSL database; melanocytic skin lesions from patients with multiple atypical nevi and with digital dermoscopy follow-up*	Dermoscopy (No algorithm) Image based (blinded)	NR; Excise	Dermatologist (n=24; exp mixed – incl basic dermoscopy experience (n=9), dermoscopy training but basic experience (n=10), experienced and trained dermatologists (n=5) Average result reported	Histology or FU MM 5, MiS 5 BN 70 10/80; 13%	None reported *10 patients with early melanomas and 10 other patients randomly selected; benign melanocytic skin lesions taken at random from these 20 participants.
Malvehy 2014 5-Referred (selected on reference) (u) MM+MiS	WPC P-CS Multicentre 1611 / 1943	Patients with skin lesions selected for total excision to rule out melanoma; dermatologists were encouraged to enrol a mix of lesions with an even distribution of low-, medium and high-risk lesions.	Dermoscopy (no algorithm; ABCD; 7PCL; 7PCLrev) Image-based (clinical image?) [Also evaluates CAD - Nevsiense]	Dx of malignancy; >4.75; >5.45; 7PCL NR	Dermatologist (n=3; exp NR) dermatologists with 2–5 years of experience in dermoscopy assessment. Unclear	Histology VI/Dermoscopy only – MM 126; MiS 112 Breakdown of non-diseased not provided for VI/dermoscopy sample [Full sample of 1942: MM 153; MiS 112; BCC 48, cSCC 1; MCC 1 BN 1497; 5 SN, 51 SK, 6 SCC in situ; 8 AK; 61 other] 238/1678; 14%	473 excluded from total sample – mainly due to investigator oversight or inability to render a final histopathological diagnosis; 74 were device-related (60 with inadequate reference measurement quality and 14 to device failure). 242 excluded from VI/Dermoscopy analysis due to image quality
Menzies 2005 4-Referred (u) MM+MiS	WPC-obs R-CS Specialist clinic Multicentre Australia, US, Germany NR / 786*	PSL imaged using SolarScan at 9 different clinical centres including specialist referral centres and private skin cancer clinics;	Dermoscopy (no algorithm) Image-based (clinical image and pt history provided) [also evaluated CAD Dermoscopy]	Subjective impression (dx of MM)	Dermatologist; (n=3 international experts); dermatolgosts (n=4); dermatology registrars (n=3); GPs (n=3) Average reported per group	Histology or FU (26% of full sample FU; 3% Expert dx) Sydney Melanoma Unit only (n=78) MM 5;MiS 6; LM 2 BN 65 13/78; 17%	*Only the 78 lesions from the Sydney Melanoma Unit included in the VI/Dermoscopy evaluation
Menzies 2008 5-Referred (selected on reference) (u) MM+MiS	WPC-algs CCS Multicentre NR / 497	Dermoscopic amelanotic (with no melanin pigmentation) or hypomelanotic (a melanin pigmentation area of less than 25% of the total surface area or slightly pigmented but with no dark	Dermoscopy (7PCL; Menzies; 3PCL) Image-based (blinding NR)	>=3; standard threshold; >=2	Dermatologist (assumed) (n=12; exp in dermoscopic evaluation scored 99 individual morphological features	Histology and FU (no.s NR; some nevi included that showed no changes	None reported * All melanomas included, and a random selection of melanocytic and nonmelanocytic lesions on

		brown, deep blue, or black pigmentation) lesions*	[also developed new algorithm on 80% of sample and tested on 20% but numbers D+/D- NR for the test set to allow 2x2 to be estimated.]		in approximately equal sample sizes. Single observer	following consecutive digital monitoring) MM 91; 14 MiS; 126 BCC; 4 cSCC BN 159; SN 11; SK 22; DF 17; BD 7; KA 1; AK 8; other 37 105/497; 21%	a non-melanoma to melanoma ratio of 3:1.
Pagnanelli 2003 5-Referred (selected on reference) (u) MM+MiS	WPC-algs R-NR Setting NR Italy NR / 20	Images of PSL from the training set of the Consensus Net Meeting on Dermoscopy (CNMD), selected by two experts*	Dermoscopy (pattern analysis; Menzies; 7PCL; ABCD) Image-based (clinical image)	Subjective impression; correct diagnosis; algorithm NR	Mixed - sec (n=16; exp NR) Average result	Histology MEL 6; BCC: 2 SK 2 ; CN 8; SN 2 6/20; 30%	None reported Data not disaggregated for BCC *pre- and post- dermoscopy training data presented for each algorithm
Piccolo 2002 5-Referred (selected on reference) (u) MM+MiS	WPC-obs R-CS Secondary Italy 289 / 341	PSL excised because of equivocal dermoscopic findings or at the patient's request	Dermoscopy (no algorithm) Image-based (clinical image) [Also evaluates CAD dermoscopy]	NR; dx of MM	Dermatologist (n=1 expert); Dermatology resident (n=1) Single observer	Histology MEL 13 SK 3; BN 316; DF 7; angiomas 2 13/341; 4	None reported
Piccolo 2014 5-Referred (selected on reference) (u) MM+MiS	WPC-obs R-CS Secondary Italy 165 / 165	Dermoscopically atypical PSL *	Dermoscopy (ABCD) Image-based (blinded) [Also evaluates CAD dermoscopy]	>4.74	Dermatologists (n=3; 1 expert, 2 non expert); GP (n=1; underwent dermoscopic training by studying an interactive atlas of dermoscopy between time periods T0 and T1). Single (results per observer)	Histology MM 23; MiS 10 BN 105; CN; 19 SN; 5 blue nevi; 3 dermal nevi. 33/165; 20%	None reported *Images assessed at T0 and at 6 mos (T1)
Pizzichetta 2002 5-Referred (selected on reference) (u) MM+MiS	WPC-algs R-CS Specialist clinic Italy 123 / 129	Small (</=5mm) melanocytic skin lesions surgically excised	Dermoscopy (pattern analysis; ABCD) Image-based (blinded)	Dx of MM; >4.75; >5.45	Dermatologist (assumed) (n=2; exp NR) Single observer	Histology MEL 5 lesions BN 124 lesions 5/129; 4%	None reported
Pizzichetta 2004 5-Referred (selected on reference) (u) MM+MiS	WPC R-CS US/Italy Secondary 151 / 151	Clinical and/or dermoscopic hypomelanotic (extent of pigmentation <=30%) and amelanotic skin lesions	1. VI (no algorithm) 2. Dermoscopy (pattern) Image-based (clinical image)	Subjective impression	NR (presume dermatologist; n=1) Exp NR; single obs	Histology AHM 34, MiS 5 BCC 25, SCC 5 BN 47, SN 5, SK 8, Other 18 39/108; 36% (analysed)	23 lesions excluded due to image quality; further 43 lesions were not available for evaluation by clinical images ("mainly benign melanocytic lesions".
Pupelli 2013 5*-Equivocal (selected on reference) (c) MM+MiS	WPC CCS Specialist clinic Italy 96 / 96	Melanomas <5 mm consecutively excised; plus 3 histologically proven small-diameter naevi per included melanoma	Dermoscopy (7-point) Image-based (RCM, site, age) [also evaluated RCM]	>=3; dx of MM	Dermatologist (assumed) (n=NR; exp NR)	Histology MM 13; MiS 11 BN 72 (incl 7 SN) 24/96; 25%	None reported
Rigel 2012 5-Referred (selected on reference) (u) MM+MiS	WPC R-NR Unclear US NR / 24	PSL analyzed as part of a prior study using a MSDSLA system (Monheit 2011)	Dermoscopy (no algorithm) Image-based (clinical image) [Also evaluates CAD Spectroscopy]	NR; Excise decision	Dermatologist (n=179; exp mixed) Average result	Histology MEL 5; 'Benign' diagnoses: 19 5/24; 21%	-
Rosendahl 2011 3–Limited testing (selected on reference) (u) MM+MiS Any	WPC-alg R-CS Aus. Primary 389 / 463	PSL submitted for histology from the primary care skin cancer practice of one author	1. VI (no algorithm) 2. Dermoscopy (pattern; chaos and clues)	1. subjective impression 2. NR; both chars present	Dermatologist (n=1) High exp (confirmed by author); Single obs	Histology MM 9; MiS 20 BCC 72; SCC 5 BN 217; BD 18; AK 14*; BNM 140 29/463; 6%	3 poor quality images excluded * AKconsidered malignant by study authors
Rubegni 2012 5*-Equivocal (selected on reference) (u) MM+MiS	WPC-algs R-CS Secondary Italy 107 / 107	Palmoplantar (acral) PSL excised over a 3 year period. All with clinical/ dermoscopic suspicious features in the absence of any clear benignity pattern	Dermoscopy (Pattern analysis; 3-step algorithm (Koga 2011)) Image-based (blinded)	dx of MM; Excise decision (3-step)	Dermatologist (n=2; exp High - 20 years' experience in dermoscopy) Single observer data	Histology MM 21; MiS 4 'Benign' diagnoses: 82 25/107; 23%	None reported
		Melanocytic skin lesions showing		NR; dx of MM			None reported

Rubegni 2016 5-Referred (selected on reference) (u) MM+MiS	NR R-CS Secondary NR-Italy 95 / 95	clear-cut dermoscopic features of regression and excised for suspected malignancy	Dermoscopy (pattern analysis) Image-based (blinding NR)		Dermatologist (n=3; exp High) experienced dermoscopists Single observer	Histology MEL 45 BN 50 45/95; 47%	
Sboner 2004 5-Referred (selected on reference) (u) MM+MiS	NC R-CS Secondary NR-Italy NR / 152	Melanocytic lesion images acquired consecutively	Dermoscopy (no algorithm) Image-based (blinded)	NR; dx of MM	Dermatologist (n=8; exp NR) Single observer	Histology ; MM 31; MiS 11 BN 110 42/152; 28%	None reported
Seidenari 1998 5-Referred (selected on reference) (u) MM+MiS	WPC-Obs CCS Secondary Italy NR / 90	Patients referred by dermatologists or general physicians with >=1 PSL difficult to interpret on clinical grounds alone, numerous PSLs, or because the patients were at increased risk for melanoma or prior malignancy	Dermoscopy (no algorithm) Image-based (blinded)	Subjective impression; dx of MM	Dermatologist (n=2; 1 expert, routinely used videomicroscopy ; 1 nonexpert) Single observer	Histology MEL 31 59 "nonmelanomas" including dysplastic nevi" 31/90; 34%	-
Seidenari 2005 5-Referred (selected on reference) (u) MM+MiS	WPC R-CS Specialist clinic Italy NR / 603	Melanocytic lesions, which had undergone surgical excision for clinical, dermoscopic, or cosmetic reasons after referral by a dermatologist for examination of a particular lesion or of the whole skin	Dermoscopy (pattern analysis) Image-based (blinded)	Correct dx of MM (atypia grade 3); Excise decision (atypia grade 2 and above)	Dermatologist (n=2) Consensus of 2	Histology MEL 112 BN 491 112/603; 19%	NR
Seidenari 2007 5-Referred (selected on reference) (u) MM+MiS	NC R-CS Setting NR Italy NR / 243	Dermoscopic images of melanocytic lesion that had undergone excision	Dermoscopy (no algorithm) Image-based (blinded)	NR; dx of MM	Mixed (n=4; exp mixed) Single observer	Histology MM 35; MiS 8 BN 200 43/243; 18%	-
Skvara 2005 7-Follow-up (u) MM+MiS	WPC-alg CCS Secondary Austria NR / 126	Consecutive lesions showing changes over time during digital dermoscopy follow-up that were excised at 2 clinics	Dermoscopy (ABCD; 7PCL) Image based (blinded)	>4.75; >=3	Dermatologist (n=2; exp High)	Histology MEL 63 BN 63 63/126; 50%	None reported
Stanganelli 1998 5-Referred (selected on reference) (u) MM+MiS Any	WPC R-CS Italy Training images Italy NR / 30	PSL images selected from computerised files of the skin cancer clinic.	1. VI (no algorithm) 2. Dermoscopy (no algorithm) Image-based (clinical image)	NR; clin dx	Dermatologists (n=20) Exp NR ("experience in ELM but (with) no formal training") Average result	Histology MEL 10 BCC 4 BN 10, SK 3, Other 3 10/30; 33%	None reported BCC results not disaggregated
Stanganelli 1999 5-Referred (selected on reference) (u) MM+MiS	WPC-Obs CCS Specialist clinic Italy NR / 30	PSL images selected from database for training study	Dermoscopy (no algorithm) Image-based (clinical image)	Correct dx MM	Dermatologist (assumed) (n=83; exp mixed) Median result pre- and post- dermoscopy training	Histology MM 10; MiS 1 14 BN; 5 BNM 11/30; 37%	None reported
Stanganelli 2005 4-Referred (u) MM+MiS	WPC R-CS Italy Specialist clinic NR / 477	Melanocytic lesions referred to Skin Cancer Unit for clinical and dermoscopic evaluation.	1. VI (no algorithm) 2. Dermoscopy (no algorithm) Image-based (Clinical image also provided) [also evaluated CAD Dermoscopy]	NR (dx of MM)	Dermatologist (n=3); GP (n=3) Dermatologists – High exp ("2 years dermoscopy experience"); exp NR for GPs, assumed Low Average reported	Histology / Registry FU MEL 31 BN 103 31/134; 23%	None reported
Stanganelli 2015 7-Follow-up (u) MM+MiS	WPC R-CS Specialist clinic Italy 70 / 70	Lesions excised on the basis of clinical and/ or dermoscopic changes at follow-up suggesting a malignancy	Dermoscopy (7-point rev - FU)) Image based (baseline image provided) [also evaluated RCM]	'major change' (dx of MM)	Dermatologist (assumed) (n=NR; exp NR)	Histology MM 11; MiS 1 BN 55; BNM 3 12/70; 17%	None reported
Stolz 1994 5*-Equivocal (selected on reference) (u) MM+MiS	NC R-CS Secondary Germany NR / 157 [79 in test set incl]	Equivocal PSLs with size smaller than 9x13 mm, melanoma tumour thickness of 1mm and melanoma Clark's level <=III	Dermoscopy (ABCD) Image-based (blinded)	>5.45; dx of MM	NR (n=1; exp NR) Single	Histology MEL: 48 (test set only) BN 31 48/79; 61%	None reported

Tan 2009 5-Referred (selected on reference) (u) MM+MiS	WPC-Obs CCS Training images UK NR / 30	Test series of images of melanomas and benign lesions	Dermoscopy (pattern analysis modified) Image-based (clinical image)	Excise decision	Mixed (n=6; exp mixed) Average result; pre- and post- dermoscopy training	Histology MEL 15 Other: 11 BN; 3 SK; 1 vascular 15/30; 50%	None reported
Tenenhaus 2010 4-Referred (u) MM+MiS	NC CCS Secondary France NR / 227	Dermoscopic images of all melanoma lesions recorded on two databases, plus 227 randomly selected benign lesions	Dermoscopy (no algorithm; based on ABCD and others) Image-based (clinical image also provided)	NR; subjective impression (dx of MM; Excise decision)	Dermatologist (n=5; exp High) Single observer	Histology plus other (65/227 benign not excised; assume expert dx) MM 28; LM 4 BN (excised) 165; 'benign' not excised: 62 32/27; 14%	None reported
Unlu 2014 5-Referred (selected on reference) (u) MM+MiS	WPC-algs R-CS Specialist clinic Turkey 115 / 115	Melanocytic lesions excised at Pigmented Lesion Clinic	1. VI (no algorithm) 2. Dermoscopy (ABCD; 7PCL; 3PCL; CASH) Image-based (blinded)	1. NR; dx of MM 2. >5.44; >=3; >=2; >=8	Dermatologist (assumed) (n=3; exp High) VI appears to be in clinic dx (single observer); derm images scored by 3 other 'expert' dermatoscopists Consensus of 3	Histology alone MEL 24 BN 91 (incl 6 SN) 24/115; 21%	None reported
Wells 2012 5-Referred (selected on reference) (u) MM+MiS	WPC CCS Industry dbase US NR / 47	PSL selected from a repository of lesions amassed during an acquisition study conducted by MELA Sciences Inc for the US Food and Drug Administration	Dermoscopy (no algorithm) Image-based (clinical image, pt history) [Also evaluates CAD spectroscopy]	NR; MM or not	Dermatologist (n=39; exp NR) Average	Histology MEL 23 'Benign' diagnoses: 24 23/47; 49%	-
Winkelmann 2016 5-Referred (selected on reference) (u) MM+MiS	WPC CCS Unclear Training images NR / 12	Selected images previously analysed by MSDSLA	1. VI (no algorithm) 2. Dermoscopy (no algorithm) Image-based (clinical image)	NR	Dermatologists (n=70) Exp NR; average	Histology MM 3; MiS 2 BN 7 5/12; 42%	None reported
Zalaudek 2006 5-Referred (selected on reference) (u) MM+MiS Any	NC R-CS Specialist clinic Italy NR / 165	Random sample of excised, equivocal and nonequivocal, PSL and and non-PSLs with melanin or haemoglobin pigmentation in all or part of the lesion.	Dermoscopy (3PCL) Image-based (age, site, gender)	>=2 chars present	Mixed (n=150; exp NR) Average result	Histology Full sample: MM 18; MiS 11 BCC: 18 79 BN; 26 SK; 8 vascular; 3 DF 26/150; 17%	15 used for training purposes 5 BCC moved from FP to to TN

FOOTNOTES: 1 Test naïve; 2 Limited prior testing; 3 Limited prior testing (with selection on reference standard); 3* Limited prior testing (with selection on reference standard and equivocal nature of lesions); 4 Referred for further assessment; 5 Referred for further assessment (with selection on reference standard); 5* Referred for further assessment (with selection on reference standard and equivocal nature); 6 Referred for further assessment (equivocal on specialist review); 7 Lesions that have been undergoing follow-up

c- clearly positioned on clinical pathway; u – unclear position on clinical pathway; NR – not reported; PSL – pigmented skin lesion; PLC – pigmented lesion clinic; MM – malignant melanoma; MiS – melanoma *in situ* (or lentigo maligna); BCC – basal cell carcinoma; cSCC – cutaneous squamous cell carcinoma; LS – lentigo simplex; SK – seborrheic keratosis; SN – Spitz nevi; AK – actinic keratosis; BN – benign naevi; BD – Bowen's disease; DF – dermatofibroma; FU – follow-up; R –retropsective; P – prospective; CS – case series; CCS – case control study; WPC – within person comparison (of tests); BPC – between person comparison (of tests); NC – non comparative; RCM – reflectance confocal microscopy; CAD – computer-assisted diagnosis

10 Forest plots of sensitivity and specificity for visual inspection for the detection of invasive melanoma and intraepidermal melanocytic variants (MM+MiS)

Figure 25

11 Summary study details for detection of invasive melanoma

Study author **Outcomes**	**Study type** **Country** **Setting** **Pts / lesions**	**Inclusion criteria**	**Index tests (algorithm)** **Diagnostic approach**	**Threshold**	**Observer qual. (n)** **Experience**	**Reference standard** **Final diagnoses** **Prevalence (MM)**	**Exclusions** **Comments**
In-person evaluations							
Ascierto 2010 MM	WPC P-CS Specialist clinic Italy 54 / 54	Clinically relevant cutaneous pigmented lesions, undergoing dermoscopy and excision	Dermoscopy (Risk stratification; modified Kenet et al) In person [Also evaluates	Very high risk; high or very high risk; correct dx of MM	Dermatologist (n=NR; exp High) Unclear observer interpretation	Histology MM 12 'Benign' 42 12/42; 22%	-

			CAD Spectroscopy]				
Coras 2003 MM	WPC NR-CS Private Germany NR / 45	PSLs undergoing excision due to diagnosis of melanoma or atypical nevus, to rule out melanoma or at the patient's request	Dermoscopy (No details; diagnosis based on clinical exam, dermoscopy, medical history) In person [Also evaluates Telederm assessment of clin/derm images]	Not reported; Correct dx of MM	Dermatologist (n=3; exp High) participating experts with great experience in dermatoscopy Single observer	Histology MM 16; 'Benign': 29 16/45; 36%	10 excluded due to poor image quality; 45 did not undergo excision
Feldmann 1998 MM MM+MiS	NC P-CS Secondary Austria NR / 500	Melanocytic lesions examined by dermatoscopy prior to excision	Dermoscopy (ABCD) In person	>5.45; >4.2	NR(n=NR; exp NR) Unclear obs interp	Histology MM 25; MiS 5 BN 272; dysplasia 190; lentigines 7; lentigo nevi 1 30/500; 6%	NR
Krahn 1998 MM	WPC P-CS Secondary Germany 80 / 80	Excised pigmented skin lesions	1. VI (no algorithm) 2. Dermoscopy (no algorithm) In person	NR; clinical dx of MM	Dermatologist (assumed) (n=1; exp NR) Single observer	Histology MM 39 BN 37; Dysplastic 2; SN 1 39/80; 49%	*none*
Piccolo 2000 MM	NC NR-CS MulticentreAustria 40 / 43	PSLs selected because of their diagnostic difficulty	Dermoscopy (no algorithm) In person observer [Also evaluates Telederm assessment of clin/derm images]	Not reported; correct dx of MM	Dermatologists (n=1; exp High) Single observer	Histology MM 11; BCC 3 SK 2; BN 23; Other 4 11/43; 26%	NR; Poor quality index test image the digital images, were assigned an image quality rating (1, excellent; 2, good; 3, sufficient; 4, poor). All images scoring 4 were excluded from the study.
Viglizzo 2004 MM	WPC NR-CS Specialist clinic Italy NR / 79	PSLs examined at the Dermoscopy Service and undergoing excisions; high and medium risk on dermoscopy were selected for excision and 2x2 can be estimated only for melanocytic subgroup.	1. VI (no algorithm) 2. Dermoscopy (no algorithm) In person	NR; correct dx of MM	Dermatologist (assumed) (n=NR; exp NR) Single observer	Histology MM 12 MN: 67 12/67; 18%	*none*
Image-based							
Arevalo 2008 MM	NC RP-CS Specialist clinic Australia NR / 3367	Melanocytic lesions imaged at the Sydney Melanoma Unit with a histopathologic diagnosis or that remained unchanged following short-term (2.5-4.5 months) digital monitoring (diagnosed as benign)	Dermoscopy (Menzies criteria) Image based (blinded)	Absence of negative chars plus >=1 positive char present; correct dx of MM	Dermatologist (assumed) (n=2; exp NR) Consensus of 2; referral to a 3rd observer if disagreement.	Histology or FU MM 341 'Benign' 3026 341/3367; 10%	*none*
Friedman 2008 MM MM+MiS	WPC CCS Secondary/Private US 94 / 99	An industry database of images of PSL <=6mm was used to sample images of melanoma and non melanoma lesions; high-grade dysplastic nevi were excluded.	Dermoscopy (no algorithm) Image-based (site, age, gender) [Also evaluates CAD-Spectroscopy]	Correct dx; excise decision	Mixed - sec (n=10; exp High) Average result (reports mean and median; mean used)	Histology MM 21; MiS 28; BCC: BN 34; SK 2; 14 other benign 21/99; 21%	None reported
Hauschild 2014 MM MM+MiS	WPC CCS Secondary/Private US 130 / 130	Subset of PSL evaluated in a MelaFind study (Monheit 2011); 65 melanoma and 65 non-melanoma randomly selected. Excluded ulcerated, non-pigmented, or located on excluded anatomic sites.	Dermoscopy (no algorithm) [Also evaluates CAD spectroscopy] Image-based (clinical image, pt history)	NR; excise decision	Dermatologist (n=101; exp High) Single observer	Histology MM 36; MiS 29 'Benign' diagnoses: 65 36/130; 28%	-
Kreusch 1992 MM	NC RP-CS Secondary Germany Full sample: 858 / 1506 (265 melanocytic included)	Pigmented lesions suspected to be malignant melanoma with adequate photo-documentation and histology results	Dermoscopy (from Kreusch 1991) Image based (slides labelled only with patient code and	>=9; correct dx of MM	Dermatologist (assumed) (n=1; 'experienced')[also presents results for inexperienced student – data not included] Single observer	Histology MM 96; BN 169	*52 NML excluded from second step evaluation*

			lesion localisation)				
Lorentzen 1999 MM	WPC P-CS Specialist clinic Denmark 232 / 232	Patients with lesions suspicious for CMM referred to outpatients clinic	1. VI (no algorithm) 2. Dermoscopy (no algorithm) Image based (clinical image)	subjective impression; correct dx of MM	Dermatologist (n=4; exp High) Average	Histology MM 49; BCC 16 SK 12; BN 137 Other: 18 (SN, BD plus others) 49/232; 21%	Poor quality index test image 10 cases excluded
Lorentzen 2000 MM	WPC-alg RP-CS Specialist clinic Denmark 258 / 258	PSL from patients consecutively referred to the skin cancer outpatient clinic with available clinical photographs, dermatophotographs and a subsequent excision biopsy were included	Dermoscopy (ABCD; Kenet risk stratification) Image based (clinical image)	>4.75; Kenet – probable melanoma; possible/probable melanoma	Dermatologist (n=3; exp High; 3 senior dermatologists with >5 y daily experience in dermatoscopy Single observer (reported per observer)	Histology MM 64; BCC 25 SK 14; BN 135; Dysplastic 3; Other: 16 64/258; 25%	-
Lorentzen 2008 MM	WPC NR-CS Specialist clinic Denmark 119 / 119	Patients referred to the specialist naevus clinic; compared classic dermoscopy to acrylic globe magnifer	Dermoscopy (Kenet risk stratification) Image based (blinded)	NR	Dermatologist (n=NR) Average	Histology MM 24; BCC 13 BN 69; Mild//moderate dysplasia 2; SK 9; Other 2 24/119; 20%	1 dermatofibroma
Menzies 1996 MM	NC RP-Unclear Image libraries Multicentre NR / 385	PSL from the Sydney Melanoma Unit with dermoscopic images and histological diagnoses; melanomas and randomly selected clinically atypical nonmelanoma lesions were included.	Dermoscopy (Menzies criteria) Image based (blinded)	2 chars absent and >=1 char present; correct dx of MM	Dermatologist (assumed) (n=NR; exp NR) NR	Histology MM 107; BCC: 18 SK 23; acquired BN 58; Dysplastic 105; Blue nevi 11; Ephelis/lentigo 17; SN 6; spindle cell nevus 2; DF 2; hemangioma 13; solar keratosis 9; other 14 107/385; 28%	-
Menzies 2013 MM	WPC-algs CCS Secondary Mixed NR / 467	Nodular malignant melanoma* and a random selection of non-nodular invasive primary melanoma, benign nodular melanocytic lesions, and nodular nonmelanocytic lesions at a ratio of NM to other subgroups of 1:2. Nodular benign melanocytic lesions and nodular nonmelanocytic lesions were identified by the clinical appearance of a solitary nodule and confirmed using dermoscopic examination.	Dermoscopy (ABCD; Menzies, CASH; 7PCL; 3PCL; Menzies (amelanotic)) Image based (NR)	ABCD >5.45; CASH >8; Menzies amelanotic >0 and >1; others at standard thresholds	Dermatologist (n=1; exp NR) Twelve scorers blinded to the lesion diagnosis scored 99 individual features in each lesion of approximately equal sample sizes, as previously described.7Following the review of the article for publication, an additional feature (blue-black structures) was scored for all lesions by one observer (E.C.). Single observer	Histology or FU ('some' benign melanocytic nevi showed no change over time compared with baseline photographs). NM 83; 134 MM BN 115; 217/332; 65%	135 NML excluded from second step evaluation *an invasive melanoma without an in situ (junctional) component beyond 3 rete ridges of the dermal invasive component
Nilles 1994 MM	NC RP-CS Secondary Germany NR / 209	Melanocytic skin lesions that underwent excision	Dermoscopy (New algorithm) Image based (blinded)	Any characteristic present?; correct dx of MM	Dermatologist (assumed) (n=1; exp NR) S ingle observer	Histology MM 41 BN168 41/209; 20%	-260 lesions used to identify best model; se/sp for overall diagnosis reported for 209 lesions investigated in 1990
Rao 1997 MM	WPC-Obs RP-CS Specialist clinic US 63 / 72	Patients with atypical melanocytic lesions or suspected early malignant melanoma	1. VI (no algorithm) 2. Dermoscopy (no algorithm) Image based (clinical image)	subjective impression; dx	Dermatologist (n=2); Melanoma Fellow (n=2) Single observer	Histology MM 21 Atypical MN 51 21/72; 29%	*none*
Troyanova 2003 MM	WPC-tests CCS Specialist clinic Not reported NR / 50	Images of PSLs selected for a dermoscopy training study	1. VI (no algorithm) 2. Dermoscopy (no algorithm) Image based (blinded)	NR; correct dx of MM	Dermatologist (n=32; exp High) Average	Histology MM 25 'Benign' 50 25/50; 50%	*NR*
Westerhoff 2000 MM	WPC-Obs CCS Specialist clinic Australia NR / 100	Clinically atypical pigmented skin lesions randomly selected from PSL image database.	1. VI (no algorithm) 2. Dermoscopy (no algorithm; Menzies criteria)	NR; dx of MM	GPs (n=74; no formal training in dermoscopy, randomised to dermoscopy education intervention (n=37) or not (n=37). Average reported	Histology or FU MM 50 'Benign' 50 50/100; 50%	*Diagnoses recorded for both groups of GPs at baseline (pre-test) and after training of one arm (post-test); post-test data for

			Image based (blinded)				the intervention group of GPs was used for the Visual Inspection analysis

NR – not reported; PSL – pigmented skin lesion; PLC – pigmented lesion clinic; MM – malignant melanoma; MiS – melanoma *in situ* (or lentigo maligna); BCC – basal cell carcinoma; cSCC – cutaneous squamous cell carcinoma; LS – lentigo simplex; SK – seborrheic keratosis; SN – Spitz nevi; AK – actinic keratosis; BN – benign naevi; BD – Bowen's disease; DF – dermatofibroma; FU – follow-up; R –retropsective; P – prospective; CS – case series; CCS – case control study; WPC – within person comparison (of tests); BPC – between person comparison (of tests); NC – non comparative; RCM – reflectance confocal microscopy; CAD – computer-assisted diagnosis; Obs - observer; alg - algorithm

12 Forest plots of sensitivity and specificity for visual inspection and for visual inspection plus dermoscopy for the detection of invasive melanoma (MM)

Figure 26; Figure 27

13 Summary study details for detection of any skin lesion requiring excision

Study author Outcomes reported	Study type Country Setting	Inclusion criteria	Index tests (algorithm) Diagnostic approach	Threshold	Observer qual. (n) Experience	Reference standard Final diagnoses Prevalence (Any)	Exclusions Comments (marked *)
In-person evaluations							
Argenziano 2006 Any	RCT Italy, Spain Primary NR / 85	Patients asking for screening or exhibiting one or more skin tumours as seen during routine physical examination (patient-finding screening). Participating PCPs randomised to either visual inspection alone or visual inspection plus dermoscopy; only excised lesions can be included for each arm.	VI (ABCD) Dermoscopy (3-point checklist) In person (single observer)	Subjective impression ; dx of malignancy	GPs (n=37) All trained in ABCD rule	Histology MEL 6 BCC 37; SCC 10 Benign 32 53/85; 62%	*Only those patients who were considered to have lesions suggestive of skin cancer had histology and could be included; rest had expert diagnosis (making full dataset ineligible for this review)
Durdu 2011 MM+MiS Any	WPC P-CS Secondary Turkey 176/200	PSL that could not be diagnosed with only dermatologic physical examination; 2x2 included for melanocytic subset	Dermoscopy (ABCD; nonmelanocytic excluded first) [Also evaluated exfoliative cytology] In person	NR	Dermatologist (n=1; exp NR) Single observer	Histology MM+MiS 10; BCC: 34; Other malignant 2 SK 24; BN 100; DF 12; Warts 16; Dirt 1; Other 1 10/200; 5%	-
Soyer 2004 Any	NC R-CS Specialist unit Italy 225/231	Lesions at pigmented lesion clinic considered by experienced dermatologists to merit excision on clinical grounds	Dermoscopy (no algorithm) In person	NR	Dermatologist (n=1; exp High)* Single	Histology MM+MiS 68; BCC 9 'Benign' 154 77/154; 33%	*Also reports data for 6 inexperienced observers interpretation of the acquired dermoscopic images; data excluded as includes 3 medical students
Stanganelli 2000 MM+MiS Any	WPC R-CS Italy Specialist clinic NR / 3372	PSL referred by dermatologists and general practitioners either for pre-surgical assessment or consultation	VI (ABCD) Dermoscopy (no algorithm) In person (Single)	NR Subjective impression	NR (assumed dermatologist - described as one of the co-authors; n=1)	Histology / Registry FU MEL 55 BCC 43; Benign 3274 98/3372; 3%	None reported
Image-based evaluations							
Carli 2002b MM+MiS Any	WPC R-CS Italy Secondary NR / 57	Clinically suspicious or equivocal PSL undergoing excision for diagnostic purposes; all <= 14mm diameter	1. VI (NR) 2. Dermoscopy (NR) Image-based (blinded)	NR	Dermatologists (n=2) High exp ('with experience in the field of '); consensus of 2	Histology MM 6, MiS 5 BCC 10 BN 31, SK 1; Other 4 11/57; 19%	4 'not evaluables' excluded (NB these differ between clinical images and dermoscopic images (1 MM excluded from VI analysis)
Lorentzen 2008 MM	WPC NR-CS Specialist clinic Denmark 119 / 119	Patients referred to the specialist naevus clinic; compared classic dermoscopy to acrylic globe magnifer	Dermoscopy (Kenet risk stratification) Image based (blinded)	NR	Dermatologist (n=NR) Average	Histology MM 24; BCC 13 BN 69; Mild//moderate dysplasia 2; SK 9; Other 2 24/119; 20%	1 dermatofibroma
Rosendahl 2011 MM+MiS Any	WPC-algs R-CS Aus. Primary 389 / 463	PSL submitted for histology from the primary care skin cancer practice of one author	1. VI (no algorithm) 2. Dermoscopy (pattern; chaos and clues)	1. subjective impression 2. NR; both chars present	Dermatologist (n=1) High exp (confirmed by author); Single obs	Histology MM 9; MiS 20 BCC 72; SCC 5 BN 217; BD 18; AK 14*; BNM 140 * considered malignant by study authors	3 poor quality images excluded

						29/463; 6%	
Stanganelli 1998 MM+MiS Any	WPC R-CS Italy Training images Italy NR / 30	PSL images selected from computerised files of the skin cancer clinic.	1. VI (no algorithm) 2. Dermoscopy (no algorithm) Image-based (clinical image)	NR; clin dx	Dermatologists (n=20) Exp NR ("experience in ELM but (with) no formal training") Average result	Histology MEL 10 BCC 4 BN 10, SK 3, Other 3 10/30; 33%	None reported BCC results not disaggregated
Zalaudek 2006 MM+MiS Any	NC R-CS Specialist clinic Italy NR / 165	Random sample of excised, equivocal and nonequivocal, PSL and and non-PSLs with melanin or haemoglobin pigmentation in all or part of the lesion.	Dermoscopy (3PCL) Image-based (age, site, gender)	>=2 chars present	Mixed (n=150; exp NR) Average result	Histology Full sample: MM 18; MiS 11 BCC: 18 79 BN; 26 SK; 8 vascular; 3 DF 26/150; 17%	15 used for training purposes 5 BCC moved from FP to to TN

NR – not reported; PSL – pigmented skin lesion; PLC – pigmented lesion clinic; MM – malignant melanoma; MiS – melanoma *in situ* (or lentigo maligna); BCC – basal cell carcinoma; cSCC – cutaneous squamous cell carcinoma; LS – lentigo simplex; SK – seborrheic keratosis; SN – Spitz nevi; AK – actinic keratosis; BN – benign naevi; BD – Bowen's disease; DF – dermatofibroma; FU – follow-up; R –retropsective; P – prospective; CS – case series; CCS – case control study; WPC – within person comparison (of tests); BPC – between person comparison (of tests); NC – non comparative; RCM – reflectance confocal microscopy; CAD – computer-assisted diagnosis

14 Dermoscopy training interventions

Study author **Outcomes reported**	**Inclusion criteria** **No. lesions; cases** **Algorithm used** **In-person/image-based**	**Clinicians recruited for training**	**Pre-training**	**Training approach**	**Post-training**
Detection of invasive melanoma or atypical intraepidermal melanocytic variants					
Pagnanelli 2003 Pathway – unclear MM+MiS	Clinical and dermoscopic images of PSL from the training set of the Consensus Net Meeting on Dermoscopy (CNMD), selected by two experts N=20; MEL 6 Dermoscopy (pattern analysis; Menzies; 7PCL; ABCD) Image-based (clinical image)	Recruited 16 'colleagues', incl medical Students (n=3), dermatology residents (n=9) and Dermatologists (n=4). All reported limited personal experience of dermoscopy, no formal training and did not use dermoscopy in daily professional practice	After the 1h lecture at the beginning of the study, lesion images were provided on CD; participants asked to complete electronic data sheet listing criteria for diagnosing PSLs by pattern analysis and by the various algorithms and to offer a dermoscopic diagnosis for each case within 20 days	1-h lecture on · basic principles of dermoscopy, the dermoscopic · features of PSLs, · pattern analysis and the diagnostic algorithms (ABCD rule, seven-point checklist, Menzies' method). Plus a web-based tutorial (http://www.dermoscopy.org); participants requested to devote 1 h per day, 5 days per week for 2 consecutive weeks.	Post-training evaluation @5 weeks after initial evaluation Participants re-evaluated the same 20 cases, again over a 20-day period.
Piccolo 2014 Pathway – unclear MM+MiS	Dermoscopically atypical PSL N=165; MM 23; MiS 10 Dermoscopy (ABCD) Image-based (blinded) [Also evaluates CAD dermoscopy]	3 dermatologists and 1 GP scored according to number of years specializing in dermoscopy, number of pigmented skin lesions assessed by dermoscopy on a daily basis, number of relevant workshops/seminars attended, and number of authored publications on dermoscopy: highly experienced (observer 1), moderately experienced, (observers 2 and 3); and minimally experienced (observer 4).	Digital dermoscopic images assessed by each observer using ABCD at T0	Between T0 and T1, Observer 4 underwent dermoscopic training by studying an interactive atlas of dermoscopy (Argenziano 2003; appears to be same as for Pagnanelli 2003)	The same digital dermoscopic images were assessed by each of the four observers using ABCD after 6 months (T1)
Stanganelli 1999 Pathway – unclear	PSL images selected from database for training study N=30; MM 10; MiS 1 Dermoscopy (no algorithm) Image-based (clinical image)	Of 223 dermatologists who participated in one of the six workshops, 83 (37%) were reported on; average of 10 y of general experience in dermatology (range 1 - 22) with routine use of ELM by 52 individuals (conventional dermatoscope for 43 and digital equipment for 9).	Pre-training test conducted after the opening lecture of each workshop (clinical classification of PSLs). Images projected onto a screen in pairs (clinical and ELM image); classified by as CMM, MN, NML, unclassifiable or equivocal; approximately 2.5 min per lesion	Nationwide educational programme in ELM; one-day meetings and workshops (duration 4 + 2 h) held with free registration. Topics included: · clinical classification and diagnosis of PSLs · management of patients with PSLs; · basic principles of ELM; · ELM criteria · ELM diagnosis; · limitations of ELM	Same set of slides re-evaluated at the end of the workshop' slides and respective correct diagnosis were discussed only after the second test.
Tan 2009 Pathway – unclear MM+MiS	Test series of images of melanomas and benign lesions N=30; MEL 15 Dermoscopy (pattern analysis modified) Image-based (clinical image)	3 consultant dermatologists and 3 specialist registrars; none had routinely used a dermatoscope.	Assessed 30 test cards consisting of one macroscopic and one dermatoscopic image of each Lesion; printed on A4 laminated paper participants classified images as 'benign',	Participants received an online tutorial (http://www.dermatoscopy.org) teaching the MPADA (Modified Pattern Analysis Diagnostic Algorithm), which could be referred to during the study period. Also each given a dermatoscope to use in clinical practice for 10	Ten months later, the test-card questionnaire was repeated (test 2).

			'malignant' or 'not known', gave a diagnosis if known, and indicated whether they would excise the lesion.	months	
Detection of invasive melanoma alone					
Troyanova 2003 MM	Patients with atypical melanocytic lesions or suspected early malignant melanoma N=50; 1. VI (no algorithm) 2. Dermoscopy (no algorithm) Image based (clinical image)	Volunteer dermatologists (n=32); experienced in clinical diagnosis of PSLs, but had no formal training in dermoscopy. ELM qualification based on good theoretical knowledge of the literature and on personal experience by trial and errors.	50 clinical images displayed individually using slide projector; scored as melanoma or 'not-melanoma'. 50 dermoscopy slides then presented and ELM diagnoses recorded. Each image shown for 30 s; no discussion of assumed diagnosis was permitted. None of the test slides used for training.	6 h of teaching daily for two consecutive days. Training was based on presentation of several hundred slides with oral explanation of the ELM criteria.	Tests were performed in the beginning and in the end of the teaching course Same test performed with slides of 50 different PSLs
Westerhoff 2000 MM	Images of PSL selected for a dermoscopy training study N=50; 50 MM 1. VI (no algorithm) 2. Dermoscopy (no algorithm; Menzies criteria) Image based (blinded)	GPs (n=74) recruited by telephone from a list of current practitioners. Required to have no formal training in dermoscopy and did not use dermoscopy in their clinical practice Participants randomized into an education intervention group or non-education intervention group (each n . 37).	Lesions presented with the clinical photograph first, followed by dermoscopy image Participants given 4 options: melanoma; benign melanocytic lesion, benign non-melanocytic lesion, `other' (specify) Clinical diagnosis recorded prior to observation of dermoscopic image.Tests completed at participants leisure	Supplied with pictorial atlas by Menzies et al.21 and a 1-h presentation on dermoscopy that specifically reviewed the Menzies method and included a quiz with images of 25 different PSL (not used in test)	As for pre-test