



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Dermato-informatic approaches to
understanding and improving lesional diagnostic
expertise in cutaneous oncology**

Roger Benjamin Aldridge

PhD
University of Edinburgh
2017

“Clinical competence is like ‘dark matter’ in astronomy: although it makes up most of the universe of working knowledge, we understand relatively little about it. What does it really consist of? Which of its components are most important? How do people acquire it? What's the best way to measure it? And how can you tell when they have enough of it? “

Frank Davidoff (1)

Declaration

As stipulated in the thesis guidelines, I can confirm with this signed declaration that:

- (a) I have composed the thesis myself, and
- (b) the work contained is my own, with any additional work undertaken as part of a research group only included where I have made a substantial contribution and that contribution has been clearly indicated, and
- (c) confirm the work has not been submitted for any other degree or professional qualification, and
- (d) that, except where indicated throughout the thesis, any included publications are my own work.

.....

R B Aldridge

Preface

The work presented in this thesis was undertaken during the period the author was employed as the clinical research fellow for the University of Edinburgh Department of Dermatology. This three-year research fellowship was financed by the Wellcome Trust project grant “Dermofit: A cognitive prosthesis to aid focal skin lesion diagnosis” - Reference: 083928/Z/07/Z (2).

During the duration of this grant the principal duties of the author were to create, develop and maintain an ontological library of skin images, designed with an integrated database architecture that could cope with the rigorous demands required for the project’s informatics studies. In addition, the author was responsible for all clinical aspects of the collaborative investigations with the Department of Informatics that were essential to enable the design and development of the software for the project. These collaborations generated a number of solely informatics publications which are listed at the foot of this preface, but not discussed elsewhere in this thesis.

Although these collaborative projects are not the major focus of this thesis, they prompted, indeed necessitated, that the author consider how individuals identify skin lesions and how expert accurate diagnosis is achieved. Furthermore, the structured lesional library of the Dermofit project uniquely provided an opportunity to ask a number of pedagogical questions that had not previously been adequately addressed. This additional work, undertaken independently by the author, was therefore ancillary but nonetheless complementary to the focus of the semi-automated software development of the Dermofit project. This thesis primarily records the author’s exploration of this opportunity.

1) *Ballerini L, Fisher RB, Aldridge B, Rees J, 2012. A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. In: ME Celebi, G Schaefer (Eds.) Color Medical Image Analysis, Springer (London) pp. 63-86. ISBN 9789400753884.*

2) *Ballerini L, Fisher RB, Aldridge B, Rees J., 2012. Non-Melanoma skin lesion classification using colour image data in a hierarchical K-NN classifier. In: A Frangi, A Santos (Eds.) Proc. International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, IEEE pp. 358 to 361. ISBN: 978145771857.*

3) Li X, Aldridge B, Rees J, Fisher R., 2011. *Estimating the ground truth from multiple individual segmentations incorporating prior pattern analysis with application to with application to skin lesion segmentation.* In: S Wright, X Pan, M Liebling (Eds.) *Proc. International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, IEEE* pp. 1438 to 1441. ISBN: 9781424441273.

4) Zanotto M, Ballerini L, Fisher RB, Aldridge B, Rees J., 2011. *Visual Cues Do Not Improve Lesion ABC(D) Grading.* In: DJ Manning, CK Abbey (Eds.) *Proc. Medical Imaging 2011: Image Perception, Observer Performance, and Technology Assessment, SPIE* vol. 7966. Bellingham: SPIE. pp. 79660U-1 to 79660U-10. ISBN: 9780819485083.

5) Li X, Aldridge B, Rees J, Fisher R., 2010. *Estimating the ground truth from multiple individual segmentations with application to skin lesion segmentation.* In: AH Bhalerao, NM Rajpoot (Eds.) *Proc. Medical Image Understanding and Analysis (MIUA), BMVA Press.* pp. 101 to 106. ISBN: 9780956615008.

6) Laskaris N, Ballerini L, Fisher RB, Aldridge B, Rees J., 2010. *Fuzzy description of skin lesions.* In: DJ Manning, CK Abbey (Eds.) *Proc. Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment, SPIE* vol. 7627. Bellingham: SPIE. pp 762717-1 to 762717-10. ISBN: 9780819480286.

7) Ballerini L, Li X, Fisher RB, Aldridge B, Rees J., 2010. *Content-Based Image Retrieval of Skin Lesions by Evolutionary Feature Synthesis.* In: C di Chio, S Cagnoni, C Cotta, M Ebner, A Ekart, AI Esparcia-Alcazar, CK Gog, JJ Merelo, F Neri, M Preuss, J Togelius, GN Yannakakis (Eds.) *Proc. Applications of Evolutionary Computation, LNCS* vol. 6024. Berlin: Springer. pp 312 to 319. ISBN: 9783642122385.

8) Li X, Aldridge B, Ballerini L, Fisher B, Rees J., 2009. *Depth data improves skin lesion segmentation.* In: GZ Yang, D Hawkes, D Rueckert, A Noble, C Taylor (Eds.) *Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), LNCS* vol. 5762. Berlin: Springer. pp 1100 to 1107. ISBN: 9783642042706. PMID: 20426221

Full texts of the above informatics papers along with the other clinical publications arising from the work described within this thesis are provided in the Appendix.

Abstract

Cutaneous malignancies represent a quarter of all new cancer diagnoses in the UK. The key to reducing the tumours' associated mortality and morbidity is early diagnosis and treatment. Prompt diagnosis remains predominately a clinical skill, but relatively little investigation of the cognitive psychology underpinning expertise in this domain has been undertaken. This thesis aims to improve understanding of these processes and investigate how lesional diagnostic expertise might be enhanced. A large database of diagnostically tagged images was captured specifically for this project. A series of separate studies were undertaken to give insight into how lesional diagnosis occurs and how it can be improved. The studies highlighted that non-analytical pattern recognition (NAPR) is likely to predominate in distinguishing malignant and non-malignant skin lesions and that the widely-promoted rules advocating analytical pattern recognition (APR) are not effective for discriminating melanoma from benign pigmented lesions. The keystone to promoting the development of NAPR and thus diagnostic expertise would seem to be increasing a novice's personal library of examples with relevant feedback. Studies demonstrated that current undergraduate exposure was variable but universally sparse, so simulation by way of diagnostically tagged images was developed which showed accuracy could be improved by increased exposure. This improvement occurred in both a content specific and dose responsive manner. These studies also highlighted that the learning curves for skin lesions are not uniform. Further studies demonstrated that the choice of images had implications on the development of diagnostic expertise; suggesting it was important that these images represent clinical practice rather than "classic" examples traditionally advocated for teaching purposes. In addition, studies highlighted the potential benefit of the 3D models developed during this project. Building on the idea that a personal catalogue of relevant referent images was crucial to enhanced diagnostic accuracy, prototype software was developed to exteriorise the experts' library of examples; in the tests described novices utilising the software delivered superior accuracy than medical students on the completion of their undergraduate teaching. In summation, the work described shows that by utilising dermato-informatic approaches lesional diagnostic competence can be improved significantly.

Lay Summary

Skin cancers are the most common malignancy in the UK. The key approach to reducing deaths and suffering arising from these tumours is early detection and prompt treatment. The work presented in this thesis explores how individuals identify such skin lesions and how their diagnostic accuracy can be improved. The key findings are summarised below:

- Pattern recognition is key to individuals' identification of lesions
- Rule based approaches (although widely promoted) are not effective
- Expertise comes from seeing lots of examples with appropriate feedback
- The number of real-life skin cancers seen by students in medical school is limited
- Simulation can be used to improve students' diagnostic accuracy
- The learning curves for the different skin cancers are not uniform
- The type of images shown can vary the effect of simulation
- Software was developed to support novices' diagnosis
- Using the software novices' accuracy was higher than trained students

The studies described show that by using modern skin cancer imaging techniques (dermato-informatics approaches) skin lesion identification can be enhanced.

Acknowledgements

The author wishes to express his gratitude to the Wellcome Trust who funded the Dermofit project, under the auspices of which the work presented in this thesis was undertaken. He wishes to also record his gratitude to the help and cooperation received from Professor Robert Fisher's Computer Vision Laboratory in the School of Informatics. The author also acknowledges with gratitude the guidance and invaluable assistance of Professor Jonathan Rees who conceived the Dermofit project and has been central to its development.

Contents

Title page	1
Dark matter	2
Declaration	3
Preface	4
Abstract	6
Lay Summary	7
Acknowledgments	8
Contents	9
Chapter 1: Introduction	10
Chapter 2: Generic materials and methods	23
Chapter 3: Novices' non-analytical pattern recognition ability in distinguishing skin lesions	34
Chapter 4: Novices' analytical pattern recognition ability in distinguishing skin lesions	49
Chapter 5: The education of medical students in the diagnosis of skin lesions	64
Chapter 6: The learning curves for the diagnosis of skin lesions	79
Chapter 7: 3D or not 3D? The role of depth perception on skin lesion diagnosis	103
Chapter 8: To what extent may "classical" images inhibit the acquisition of diagnostic skills for skin lesions	111
Chapter 9: Exteriorising the experts' internal reference library for skin lesions	129
Chapter 10: Hazards of single skin lesion assessment	142
Chapter 11: Summary and Conclusions	157
Abbreviations	162
Bibliography	163
List of the publications, abstracts and presentations arising from the author's work on the Dermofit project	179
Appendix of published work	183

Chapter 1

Introduction

1.1 Skin Cancer

Cutaneous malignancies now account for at least a quarter of all new cancer diagnoses in the UK (3). In Scotland, where a predominately Caucasian skinned population lives in a temperate latitude, skin cancers are currently estimated to affect almost 20% of the population over their lifetime (4). The risks of developing such tumours rises significantly for similar Fitzpatrick skin types who reside in climates where there is greater sun exposure, such as the USA (40% lifetime incidence) and Australia (up to 70% lifetime incidence) (5-7).

The aetiology of these malignancies is complex and not a primary focus of this thesis (8,9). However, as with most human carcinogenesis, skin tumours arise as a product of the interaction of a variety of genetic influences with varying environmental factors and in particular a complex relationship to UV exposure (10-13).

The term cutaneous malignancies, in practice, refers to three main types of skin cancers, with the common malignancies being readily divided into two groups based by the primary cell of origin. Rarer skin tumours such as Merkel cell carcinoma have not been specifically considered in this thesis.

Melanomas

Melanomas arise from melanocytes, which are the pigment producing cells in basal epidermis (9). Melanomas are the least common of the three skin cancers with 1248 cases recorded in Scotland in 2014, giving a crude incidence rate of 23.3/100,000 (4). Melanomas are, however, responsible for the most deaths, with 176 deaths attributable in 2014, giving a crude standard mortality ratio of 3.3/100,000 (4). The lifetime risk of melanoma in Scotland is recorded at 1.63% (4).

The key prognostic feature in determining patient mortality is the thickness of the melanoma in the skin at the time of identification (14). This is referred to as the Breslow depth and is measured by identifying the distance between the stratum granulosum in the epidermis and the deepest invasive melanoma cells identifiable histologically (15). The inverse relationship between Breslow depth and patient survival is shown in Figure 1.1a (16).

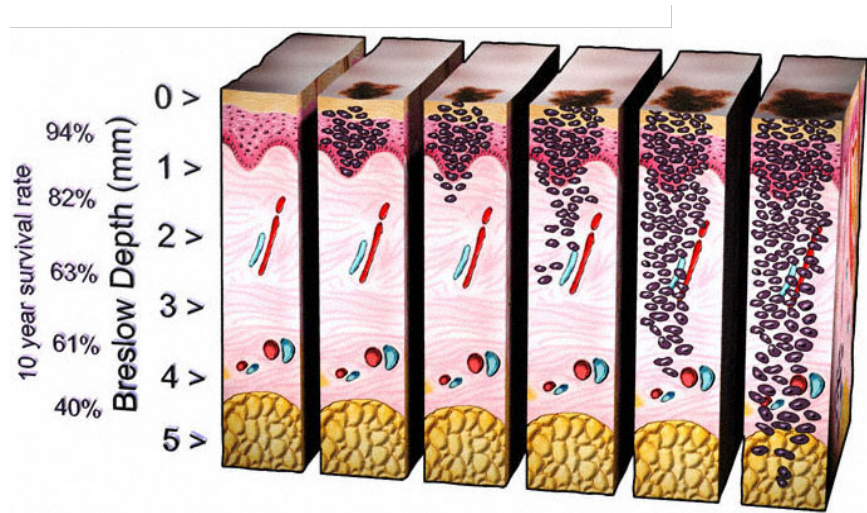


Figure 1.1a: Schematic diagram showing the relationship between increasing Breslow depth of melanoma and decreasing 10 year patient survival rates (16).

Non-melanoma skin cancers (NMSC)

NMSC arise from keratinocytes and encompass basal cell carcinomas (BCCs) and squamous cell carcinomas (SCCs) (8). Although the data for NMSC are widely acknowledged to be incomplete, the latest Scottish statistics demonstrate that BCCs are the most common of the skin cancers with 8221 cases recorded in the Scotland in 2014, giving a crude incidence of 153.7/100,000 (4). The lifetime risk of BCCs in Scotland is estimated at 11.9% (4). Whilst metastasis from BCCs is rare (17), as the tumours have a propensity for the head and neck their management can result in significant morbidity unless they are dealt with in a timely manner (18).

SCCs lie between melanomas and BCCs in terms of both incidence and mortality, with 3130 cases recorded in Scotland during 2014, giving a crude incidence rate of

58.5/100,000. The lifetime risk of SCC in Scotland is recorded at 3.4% (4). Again because of the propensity for aggressive SCCs to arise within the head and neck and metastasise to regional lymph nodes basins, early diagnosis and prompt treatment is the cornerstone to reducing morbidity and mortality (19). The mortality data is pooled for the NMSC in Scotland with a combined total of 94 deaths during 2014 and a standard mortality ratio of 1.76/100,000 (4).

Although the three tumour types are distinct pathological entities with different disease profiles, a universal theme is that the incidence of all three tumour types has been rising significantly and is predicted to continue to do so well into this century (20,21). Cancer Research UK (CRUK) statistics show melanoma incidence has quadrupled since the 1970s and increased by almost half in the last decade - see Figure 1.1b below (3). Similar trends have been also been reported by the American Cancer Society from the USA (22).

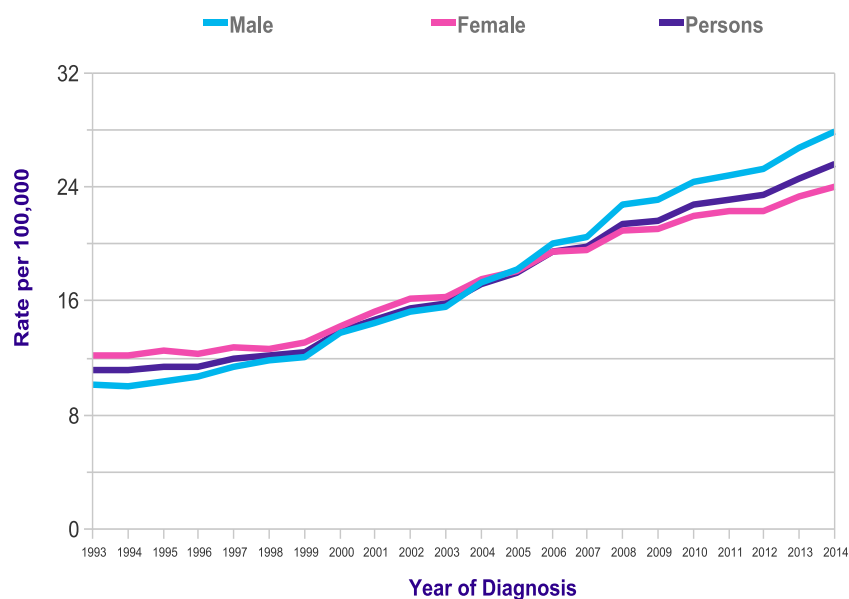


Figure 1.1b: CRUK graph demonstrating melanoma incidence rate (per 100,000) over the years 1993-2014 (3).

It is widely accepted that the origins for the significant rise in recorded skin cancer incidence are multi-factorial. The most common hypotheses advanced to explain this

upward trend in skin cancer registration figures include; an increasingly ageing population (albeit corrected for with age standardised rates), recreational changes in individuals' sun and artificial ultraviolet (UV) exposure, together with environmental changes in atmospheric UV radiation (23,24). In addition to any such "true" drivers of increased incidence there is also a potential contribution from both ascertainment bias and diagnostic drift (25,26).

Irrespective of the mixed aetiology, the result has been a change in the economic burden of managing these cancers to worldwide health services (27-31). The costs for managing cutaneous malignancies in Australia is now on a par with managing cardiovascular disease, consuming more resources than any other tumour type (32). Whilst the lifetime risk of skin cancers in Australia is 3-4 times that of the UK, even in the UK the cost of managing skin cancers has become a significant economic burden (33). The direct UK costs have been estimated at £106 million for England in 2008 with a substantial increase to £180 million being predicted by 2020 (34).

The cornerstones of reducing both morbidity and mortality (and potentially reducing the economic burden) of skin cancers is both prevention and early detection. The evidence for this has been the many studies showing that for melanoma a statistically significant relationship exists between the degree of disease involvement at diagnosis and mortality (35-37). Although intuitively it appears logical that early diagnosis and treatment of NMSC will not only reduce morbidity but also the costs of treatment, the supporting evidence is, to date, limited (38).

The cost benefits of both a preventative service or screening programme cannot solely be assessed by comparing the financial costs incurred as a result of that service with the potential medical costs saved by preventing the development of the pathological entity for which the preventative or screening measures were introduced. Some economic value must be placed on the outcomes that the patients' most care for, such as years of life saved and quality of life. When such outcomes are awarded an economic value something that is cost effective therefore may not be cost saving, and indeed very few such services save more money than the cost incurred in their prevention (39). Preventative programs for cutaneous malignancies such as Australia's

SunSmart programme have, however, been shown to be both highly cost effective and cost saving (40).

Although the cost benefits of screening (as opposed to preventative) programmes for cutaneous malignancies remain controversial, several studies have established early skin cancer detection by screening can result in lower mortality and, depending on the financial thresholds applied, melanoma screening has been suggested to be cost effective (41-45). Several studies have examined the benefits of whole body screening compared with screening lesions of concern which will be the subject of further discussion in this thesis (46,47). Although several devices for automated assessment of lesions have been developed, all currently lack the accuracy and breadth to cope with population based demands, so the early detection of skin cancers remains primarily a clinical skill (48,49).

1.2 Diagnosis of skin cancers

Clinical diagnosis of skin cancer is largely a perceptual skill, relying little on formal or explicit rules, but rather on prior exposure and feedback either in a training environment or in the clinic. Recognition of morphological patterns plays an essential role in all dermatological diagnosis and lesional dermatology is essentially spot recognition. This is illustrated in the non-dermatological example below.



Figure 1.2a: Three photographs illustrating different animal camouflaging “spots”. Images courtesy of Professor J L Rees.

To the non-expert eye, although subtle differences in the arrangements of spots may be noted, without experience useful classification is not possible. Experts, however, can readily identify these patterns and attach diagnostic labels.

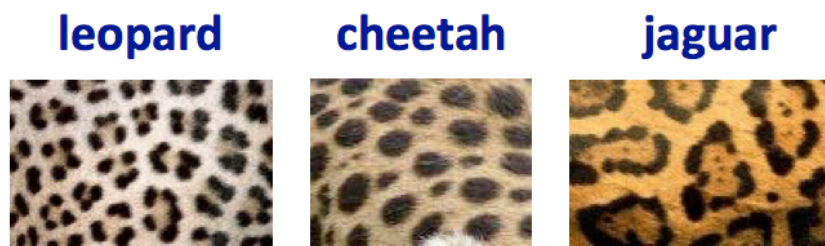


Figure 1.2b: Same three photographs from Figure 1.2a, but with “diagnostic” labels attached. Images courtesy of Professor J L Rees.

Although intellectually many dermatologists may be marginally offended by such an analogy being applied to their diagnostic skills; attaching semantics to images (i.e. being able to name, and hence categorise entities on the basis of their appearance) is the cornerstone of lesional dermatology diagnosis, which itself is the primary step in managing cutaneous malignancies. Interestingly such spot recognition is not helped by, nor does it require, background epidemiological knowledge to distinguish the possible differential diagnoses.



Figure 1.2c: Same three photographs from Figure 1.2a, but with “epidemiological” data attached. Images courtesy of Professor J L Rees.

Nor to extend the analogy further does an in-depth understanding of species classification aid the visual splitting process.

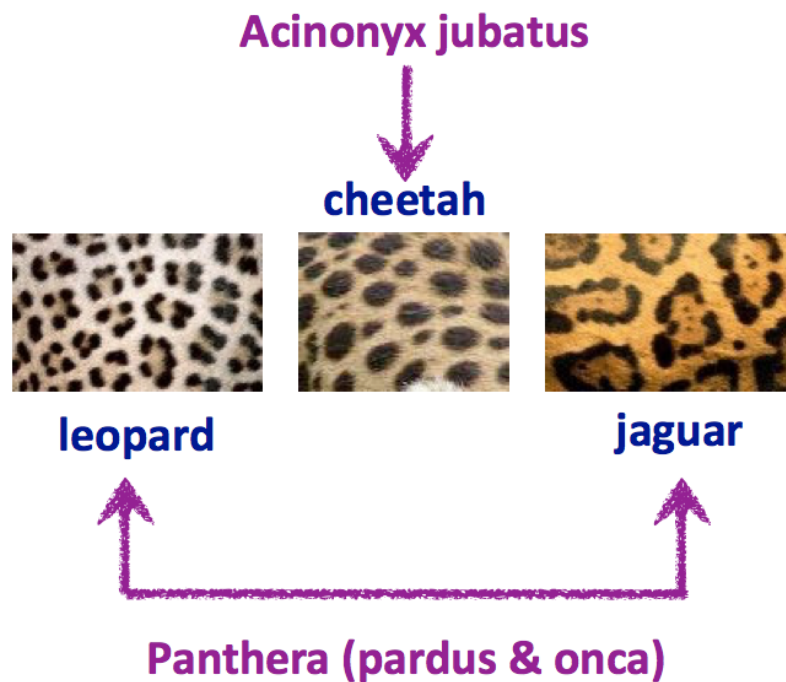


Figure 1.2d: Same three photographs from Figure 1.2a, but with biological classification attached. Images courtesy of Professor J L Rees.

So what are the processes that allow experts to diagnose skin lesions? Although mooted by Jackson in 1975, the cognitive processes involved in dermatological diagnoses were first formally investigated in a 1989 Archives of Dermatology paper by Geoff Norman (50,51). Over the last 25 years his group, amongst others, have investigated the psychological processes involved in developing expertise in dermatology and other medical specialties (52-58). At the risk of some simplification the resultant evidence suggests that for visual specialties, there are two main processes by which visible features are examined to achieve diagnostic identification.

First, analytical pattern recognition (APR), where rules are used to aid diagnosis in an explicit manner and based on conscious analytical reasoning. The ABCD rule for melanoma is such an example (59).

Second, non-analytical pattern recognition (NAPR), where identification is made without conscious thought or reference to specific rules in an implicit manner and hidden from the conscious awareness of the diagnostician. NAPR includes overall pattern recognition (OPR) where identification is achieved intrinsically as a result of previous clinical experience, and differential pattern recognition (DPR) where identification is achieved by noticing the odd one out. Grob's ugly duckling sign for melanoma being an illustration of the latter (60).

To illustrate these different recognition processes another example may be helpful. Consider the following images.



Figure 1.2e: Four photographs of James Bond cars. Clockwise from top left- Aston Martin DB5, Citroen 2CV, Aston Martin V8 Vantage and Aston Martin V12 Vanquish (61).

All readers of this thesis will have almost instantaneously identified the four images as cars. They did this not by APR, following a set of learnt rules (e.g. four tyres, steering wheel, metal chassis, headlights, windscreen) but by NAPR - in particular OPR. The non-utility of APR is highlighted in this example of everyday object recognition as many of the features that define a car could equally be applied to a truck, bus or other motor vehicles, despite which, even non-speaking toddlers are able to determine the

differences due to their previous exposure to examples with constructive feedback.

Some readers will have been able to further refine their diagnoses and identify the specific model types, or realise that they were all driven by James Bond, not through any significant knowledge of automotive engineering, but again through OPR but this time by accessing their larger personal reference library. Finally all readers will have been to identify the Citroen 2CV (the yellow car) as the odd one out - again through NAPR but this time by DPR.

Medical diagnosis is not exclusively NAPR- experts use both APR and NAPR. Schmidt suggests the experience of experts enables them to develop exemplars of disease, which he terms the "illness script". Such exemplars enable accurate and rapid diagnosis, employing a minimum number of data points, provided the disorder accords with the abstracted prototypical "illness script" pattern. Where disease is complex and does not conform to the "illness script", the expert clinician resorts to the more deliberate and time-consuming process of the analytical reasoning employed by the novice (62-64).

Irrespective of the nature of intuitive reasoning studies employing magnetic resonance imaging (MRI) suggests the actual process employed by an expert diagnostician occur in the ventral medial pre-frontal cortex (65). Patel et al opines that the errors in both novice and expert occur when diagnosis result in reasoning from hypothesis to evidence, whereas 'correct' expert diagnosis typically progresses from evidence to hypothesis (66).

Pople has suggested that experts may employ a heuristic method based on a single sign or symptom as a method of imposing order on an ill-defined presentation. In contrast with the common perception that diagnosis employing analytical reasoning is more assured than that based on intuitive reasoning Pople's hypothesis suggests that even in complex disorders the expert relies on intuitive elements in their diagnostic process (67). It would seem likely that both elements are critically interwoven, as the experts' ability to diagnose correctly is directly correlated to experience and knowledge which increasingly favours intuitive over analytical reasoning. The former enabling

an expert to cope with heavy workloads, whilst retaining or perhaps even developing an awareness of when analytical reasoning is appropriate or necessary to achieve accuracy (68).

The observation that the balance of recognition and inference in diagnosis varies with experience, and perceptual learning is acquired early necessitating emerging cognitive capabilities having to contend with a stronger existing perceptual ability, is supported by studies in radiology (69).

Dermoscopy or epiluminescence microscopy is a technique described by Goldman in 1951 (70), which has in recent years gained increasing traction amongst dermatologists for assessing cutaneous lesions (71). Although in experienced hands it has been shown to significantly improve diagnostic accuracy compared to naked eye examination (72,73), when initially introduced despite (or possibly because) dermatoscopic diagnosis was based on a series of features or rules aiming to aid the user in distinguishing malignant from benign lesions, several studies showed that the diagnostic accuracy of experienced dermatologists decreased when they initially began using a dermatoscope (74,75). Although such rules are still advocated for those new to dermoscopy, it has been well established that the diagnosis of dermatoscopic experts is based on NAPR developed through experience rather than any APR based feature processing (76,77). The introduction of dermoscopy, within a cadre of experts, provides an extremely pertinent example not only of the switch of APR to NAPR with increased exposure to exemplars but also the potential hazards of a diagnosis based solely on APR.

1.3 Accuracy of skin cancer diagnosis

Experience of prostatic and thyroid cancers has revealed the existence of a large reservoir of non-aggressive cancers and some authors have suggested the same may be true for melanoma (78-81). Longitudinal studies on melanoma to establish the frequency or indeed the existence of histologically malignant but behaviorally benign melanomas are clearly unethical and a pragmatic approach to lessen the mortality of skin cancers remains early detection and removal (82). In addition to the uncertainty

over whether histologically malignant skin lesions will behave aggressively there is also well established variance in the histological diagnosis of clinically suspicious pigmented skin lesions (83,84).

Accurate clinical diagnosis, therefore, remains integral to minimising the costs and morbidity associated with the biopsy of non-malignant lesions. Many studies have attempted to determine the accuracy of specialists and non-specialists in the diagnosis of skin cancers, but the subjects, methodology and data presented vary significantly rendering a global interpretation complicated and labour intensive. To illustrate the variability Pearl (on trainee surgeons) showed a sensitivity of 100% and specificity of 95.6% for melanomas (85); Hallock (again on surgeons) reported a sensitivity of 73% and specificity of 90% for melanomas (86); but Ek (also on surgeons) claims sensitivity of 47.8% for melanomas (87). In addition, studies such as Heal's on a mixed group of specialists and non-specialists where sensitivity was 55.6% for all malignant lesions on shoulders but lower on other regions of the body suggesting accuracy of diagnosis can be site specific raise further difficulties (88).

Unsurprisingly studies have consistently shown that the more experienced diagnosticians perform better than non-experts (89-91). Such findings are corroborated in additional work that shows improved expert clinical diagnosis results in the lower removal of benign lesions (92-95). Therefore, for the same reasons, any interventions that are instrumental in improving diagnostic accuracy are likely to result in cost savings.

The current Cochrane suite of reviews on the diagnosis of skin cancer, for which the author has been a clinical contributor since the programme's commencement in June 2014, is attempting to clarify the position regarding clinical diagnostic accuracy of skin malignancies across different groups of generalists and specialists in addition to assessing the role adjuncts may have in improving their accuracy (96,97). Although the review is not yet ready for publication the provisional findings support the importance of experience in the diagnosis of skin malignancies.

1.4 Aims of Thesis

It is clear that the burden of screening large populations for skin cancer is significant and, if extended, unlikely to be achievable using the currently available pool of expert dermatologists. If our present screening process are to be extended it may need to involve initial assessment by either non-specialist medical staff or non-medical technicians - such as physicians' assistants - specialising in lesion assessment. It is therefore essential to identify whether non-experts have the same innate abilities as experts to identify skin lesions and, if so, how to improve their training to maximise their diagnostic accuracy.

A primary aim of this thesis is to explore the features leading to accurate lesional identification by non-experts. Initial studies investigated novices' NAPR and APR abilities in distinguishing malignant and non-malignant skin lesions.

Although subject to criticism, in the current UK health model GPs act as the gatekeepers to expert assessment of problematic skin lesions (98,99). As many of these GPs will receive their sole dermatological training as undergraduates the effectiveness of the education of medical students in the diagnosis of skin lesions was also investigated.

Using the findings of these preliminary studies, a secondary principle aim was to investigate how the diagnostic accuracy of non-experts could be improved. Studies were undertaken to investigate the potential of techniques designed to improve the learning curve of non-experts in the diagnosis of skin lesions. The use of both 2D and 3D images to assesses the role of depth perception on skin lesion diagnosis was examined. Assessments were undertaken of the misleading effect the use of "classical" images may have on the engendering of false confidence and the possible role of such images in inhibiting the acquisition of diagnostic competence.

A significant proportion of the work of this thesis involved the use of modern techniques of electronic image capture and cataloguing lesions. In addition to using this library of images to investigate improvements in the teaching of non-experts, a

prototype software was examined which attempted to mimic or exteriorise the internal mental reference library used by experts in their diagnosis of skin lesions.

Finally, the outcome of total body skin examination was compared to that restricted solely to index lesions and the relevance of these findings, along with the experience of others, considered in respect to the early diagnosis of cutaneous malignancies.

Chapter 2

Generic materials and methods

Study specific details about the materials and methodology will be outlined separately in each chapter. Nevertheless, to minimise repetition the following is a summary of the generic methodology applicable to all the studies presented within this thesis.

2.1 Ethical approval

The principle NRES (National Research Ethics Service) application was made by Professor Jonathan Rees as part of the process for being awarded the Wellcome Trust project grant that funded this work (Reference: 083928/Z/07/Z). The Lothian LREC 03 (Local Research Ethics Committee) approved the NRES application in May 2008 (Reference: 08/S1103/22). Authority was subsequently sought to make a number of amendments to this initial application which are summarised below.

In March 2009 amendments were sought and granted to allow:

- i) Lay volunteers to assess the collected images.
- ii) Posters highlighting the research project to be displayed in the dermatology waiting areas to improve patient recruitment and image acquisition.

Furthermore, the original NRES application had only included a single SSI (Site Specific Information) - namely the Department of Dermatology, Lauriston Building, Edinburgh. To improve overall image acquisition and allow better collection of sequential images (particularly relevant for the studies outlined in Chapters 8 & 10) it was important to extend the image collection to a second SHA (Strategic Health Authority) - namely Stobhill Hospital, Greater Glasgow & Clyde. Although the project was SSA exempt (Site Specific Assessment) the advice received from the local ethics officer was to defer applying for any site extension until after the IRAS (Integrated Research Application System) was introduced in Spring 2009. After this date a LREC substantial amendment was not required for a site extension, instead approval was only needed from both the local management and local R&D (Research &

Development) in the secondary site. On this basis, a site extension was sought and granted through IRAS in November 2009.

Medical students

To allow the studies to recruit medical students additional permission was needed from the University of Edinburgh. An application was submitted to the University's "Committee for the use of medical student volunteers" who granted approval for the medical students involvement in these research studies in May 2009 (Reference: 2009/8). No students were recruited to more than one study to avoid any potential interaction effect.

2.2 Patient recruitment

Although the majority of images were not from consecutive patients the library was designed deliberately not to be a typical departmental library of "classic cases" but to be representative of routine clinical practice. In this manner all patients who attend the dermatology department with suitable conditions (i.e. any lesions) were asked by the doctor they saw in clinic if they were willing to participate in the research study. If they expressed an interest, and were able to spare 15 minutes, they were provided with an information sheet and given time to read it. They were then seen by the author and asked if they were willing to formally consent provided they met the appropriate inclusion/exclusion criteria detailed below.

Inclusion criteria

- 1) Age >16
- 2) Able to provide meaningful consent
- 3) Lesional skin disease (i.e. focal lesions that might be mistaken for skin cancer)
- 4) Lesions on body sites not considered embarrassing
- 5) Patients referred to the dermatology clinic with 'lesions' for diagnosis and management.
- 6) Able to spare 15 minutes.

Exclusion criteria

- 1) Age<16
- 2) Patients not able to understand what is involved
- 3) Lesions on buttocks, genitals, breasts or upper thighs.

Improving recruitment

Several strategies were implemented to improve patient recruitment:

First on the days when the Departmental lesion clinics were being undertaken the author would individually speak to all the clinicians involved in assessing patients, to encourage them to refer any suitable lesions, in addition, reminder cards were placed on the clinicians' desks during these lesion clinics.

Second, after the granting of the approvals detailed above, posters informing patients of the project which stipulated how to get involved were placed in prominent positions in the waiting areas of the Department. These posters improved patient knowledge and encouraged individuals to volunteer to the doctor assessing their skin complaint if they felt they had a lesion that they felt may be suitable for including in the study.

Third, every 3-months, at the departmental weekly education and audit meeting, the author would provide an update on the research work undertaken along with statistics on patient recruitment and referral rates.

Enriching rarities

In addition to the measures described above to improve the rate of recruitment through from the dermatology clinics, as the lesional library developed it became clear that certain categories of lesions would need to be enriched to ensure that adequate examples of all key diagnoses were included within the collection. One such key diagnosis that required additional recruitment was melanomas. Their paucity

reflected the fact that only 80 cases per year came through the department. It was therefore critical not to miss the opportunity to photograph any suitable lesions in willing participants. To enrich the inclusion of melanomas further clinician education was undertaken emphasising the importance of capturing any suspect pigmented lesions. In addition, on the days of the pigmented lesion clinics the author based himself within the biopsy waiting room to actively enroll all suitable and willing patients.

Sequential images

As explained above additional ethical permission was sought to allow the extension of the project to a second SSA. There were several reasons to pursue this strategy:

First the structure of the clinic at the Stobhill Hospital, which in effect was a local district general hospital (DGH), allowed the potential photography of sequential patients over a prolonged period. The recruiting and photographing at the main University of Edinburgh Dermatology department would not have allowed all sequential patients to have been captured as the workload would have far outstripped the available capacity of the author.

Second the dermatologists involved in providing the skin lesion service in this DGH were very proactive in patient recruitment facilitating a high level of capture.

Third as the Glasgow DGH clinic occurred on a day when no lesions clinics were running in Edinburgh it added further opportunity to enrich the collection of melanomas.

Finally, as described in Chapter 10, the practice at this lesion clinic was to offer a total body skin examination (TBSE) to each patient, thereby allowing the recording of both index and incidental lesions.

2.3 Image acquisition

All the images used in the described studies were prospectively collected for this project. Unless otherwise stipulated these were captured using the same custom built photographic set-up, designed and built in conjunction with Dimensional Imaging (DI, Glasgow) to allow for simultaneous acquisition of both two dimensional (2D) digital photographs and three dimensional (3D) models of the skin lesions (100). The dual-camera fixed distance setup is demonstrated below.



Figure 2.3a: Photograph to show dual camera setup.

The cameras were a pair of identical Canon (CANON, Japan) EOS 350D 8.1MP fitted with Sigma (SIGMA, Japan) 70mm f2.8 macro lenses and a Sigma EM-140 DG Ring Flash mounted on the camera perpendicular to the photographic field. The camera fitted with the ring flash (on the left of the picture above) was considered the “master” camera as the image was at 90 degrees to the lesion being photographed. The cameras were held at a fixed distance of 50cm using an aluminium rig which was mounted on a variable height tripod. The patients’ lesion(s) were centred in the blue 10x10cm photographic field square. The blue mount had metric scale markings on three sides and a Pantone (PANTONE, New Jersey) colour chart superiorly. Posterior to the

photographic field was blue background photographic paper. The cameras were connected to a PC through dual USBs.

Image coding

Each patient's lesion(s) photographed was allocated an unique non-identifiable code, consisting of an alphabetical letter followed by a three digit number. These codes were attached to the blue photographic field with a handwritten label. The unique codes were linked to the patients' consent forms and identifiable patient details for only so long as it was required to confirm the histological diagnosis (as outlined below under data protection). Thereafter the images were entirely anonymised.

2D images

Every time the shutter release button was depressed on the master camera a pair of digital photographs were taken - one from each of the two cameras. These were captured in the cameras' RAW + L (Large Fine) mode which simultaneously saved both RAW + JPEG files in 3456x2304 resolution. As a result, from each photograph four digital images were generated. For the purposes of the 2D studies outlined in this thesis the JPEG from the master camera was used.

3D models

For each of the patient's skin lesion(s) 3D models were generated using Dimensional Imaging's 3D proprietary software. This software employs passive stereo photogrammetry to capture the 3D surface features of skin lesions. This technology allows extremely dense 3D models to be recovered simply by using two cameras to take images simultaneously of any given lesion, from two different angles. The software constructs a 3D model of the lesion using an algorithm based on triangulation and other than the time required to build and store the models the procedure is nearly as simple as taking an ordinary clinical digital photograph. Once the 3D image is generated it can be 'zoomed in' and 'panned around' by the user to

be viewed from different angles on any PC, just as would be possible with a real-life or prosthetic example (See Chapter 7).

Sequential images

As the custom-built rig was not transportable to Glasgow on a weekly basis the images from the sequential patients recruited in Chapter 10 were acquired using a slightly different camera setup. All the lesions photographed in Glasgow employed the same photographic setup camera; a Canon (CANON, Japan) EOS 5D Mk1 13.1 MP digital camera, Canon EF 100mm f2.8L Macro IS lens and Canon 5EX ring flash. The photographs were taken at a set distance directly perpendicular to the skin lesion. The images were labelled in the same manner with a unique alpha-numerical code. In addition, each photograph had a stuck on metric scale to correspond to those taken on the rig illustrated. Given that these were single photographs that lacked triangulation it was not possible to build 3D models for these lesions.

2.4 Database construction and image cataloguing

Image cataloguing

Once all the images from the cameras were downloaded to the PC the 3D models were built using Dimensional Imaging's software (DI, Glasgow). As this software was Windows based (MICROSOFT, California) this stage had to be undertaken on the PC. Once the 3D models were built all the files were transferred to a iMAC (APPLE, California). At this stage the files were renamed according to the unique alpha-numerical code given to each of the patients and organised in a hierarchical folder structure. Each patient episode therefore had a folder named by their alpha-numerical code that contained all the relevant 2D digital photographs along with the associated 3D models. In addition to the full sized digital photos outlined above (that were used for research purposes) additional 'thumbnail' images for each of the lesions were generated by resizing and compressing all the full-scale images in Photoshop (ADOBE SYSTEMS, California). These thumbnails were of the size (approximately 10% of

original file) that then allowed them to be directly integrated into the database created (see section below).

Database construction

The Dermofit lesion database was constructed using Bento v2 (APPLE, California). Bento was a database application for Mac OS X based on FileMaker Pro (APPLE, California). It allowed a customisable structure to define the fields required for the project- see Figure 2.4a below. These fields included; the anonymised patients' demographic details, details regarding image acquisition, the diagnoses of the images (clinical, pathological, image and final), diagnosis subtype (including the author's visual sub-classification), technical details regarding the 3D models and quality of the 2D photographs, along with the author's subjective measures of lesion typicality and the lesions image thumbnail. Where discrepancies arose between the clinical, pathological and image diagnoses the final diagnosis was agreed by a dermatological Dermofit project team consensus review (BA & JLR).

Figure 2.4a: Screenshot of the Bento “Dermofit” lesional library.

The database was set up with check-boxes to highlight any diagnostic discrepancies and confirm when the images meet the technical criteria for inclusion in the collaborative work with the school of informatics. When a clinical diagnostic discrepancy was identified, this was resolved by further assessment of the lesion by a second clinician (Professor J L Rees). The structure of the Bento database allowed exporting of the data into Excel (MICROSOFT, California) in tabulated form for research purposes.

Data protection

The final digital images and 3D models are anonymous. No patient identifiable material was required to physically leave the Department of Dermatology. Patient identifiable material (linked to the images using the coding method described above) was only required for matching of the image to the correct histological diagnosis, which was normally accomplished within 4 to 12 weeks. Once this was performed, the reference link to the identifiable material was destroyed, with the only information kept being the anonymous image. Whilst identifiable material was kept (name, date of birth and patient number), it was stored on a separate non-internet connected computer, in an encrypted vault with password protection. The computer in turn was password protected and kept behind a locked door.

The paper consent forms were kept for a period of three years and then destroyed. Whilst in storage these consent forms were stored in a locked filing cabinet, again, behind a locked door. The database that keeps identifiable material (on a temporary basis) was backed up to an encrypted disc image, which was then kept within the School of Informatics.

Data backup

The importance of digital data management including backup strategies in academia is well reported (101). This was especially true for the Dermofit project, where the entire success of the project hinged on the digital data generated for the collaborative dermato-informatics approaches. As a result, the author designed and implemented a

strict back-up protocol for the Dermofit library and allied digital data. In brief this consisted of; setting up automatic backups with Time Machine (APPLE, California) on the main project computers, with additional weekly full library backups to external hard drives and monthly off site encrypted backups to dual sites.

Database size

The Dermofit library had over 5000 clinical images catalogued by the project end date. A significant proportion of these have now been made available to other investigators for informatics research through the Edinburgh Dermofit Library (102).

2.5 Undergraduate teaching

The University of Edinburgh's standard undergraduate dermatology teaching program is similar to the majority of other UK Medical Schools (103,104); consisting of an introductory series of lectures, followed by clinical exposure in outpatients. In Edinburgh, there are nine lectures and ten half-day clinical sessions over a two-week attachment undertaken in the student's penultimate clinical year. The clinical sessions incorporate eight small group demonstration clinics where the instructor has no clinical responsibility and where patients are recruited from up to ten adjacent NHS feeder clinics, a "one-on-one" NHS clinic with a consultant and a skin surgery session. The whole attachment is undertaken in the regional teaching hospital, with a departmental throughput of in excess of 25,000 new patients per annum (105).

Any additional tuition or exposure to images is specifically detailed in the material & methods sections of the individual studies and was undertaken in addition to the standard attachment outlined above.

2.6 Data analysis & Statistics

In recording the data the following techniques were employed:

- i) Paper based, purpose built, free text answer sheets or paper based, multiple choice answer sheets were used for the majority of the studies which investigated non-expert diagnosis.
- ii) For some of the studies, where only a limited number of multiple choice answers were required, it was possible to collate the subjects' responses automatically using Interwriters' audience participation radio-frequency based Crickets with their Response software (v6.70) (106).
- iii) For the studies where specific software had been designed in conjunction with the department of informatics, a data recording facility was integrated as part of the coding for these software programs.

Irrespective of the method of collation, all data was tabulated into Excel spreadsheets (Microsoft for Mac, California), from which it was possible to export it as a Tab delimited text file (.txt). The text files were then imported into R for statistical analysis and graphics generation (107). All programming in R was undertaken by the author with additional verbal guidance when necessary from Professor J L Rees. Statistical significance and associated P values are stated independently for each of the studies. Where stated random number generation was used for selecting images. These random numbers were computer generated using the R command "sample".

Chapter 3

Novices' non-analytical pattern recognition ability in distinguishing skin lesions

3.1 Introduction

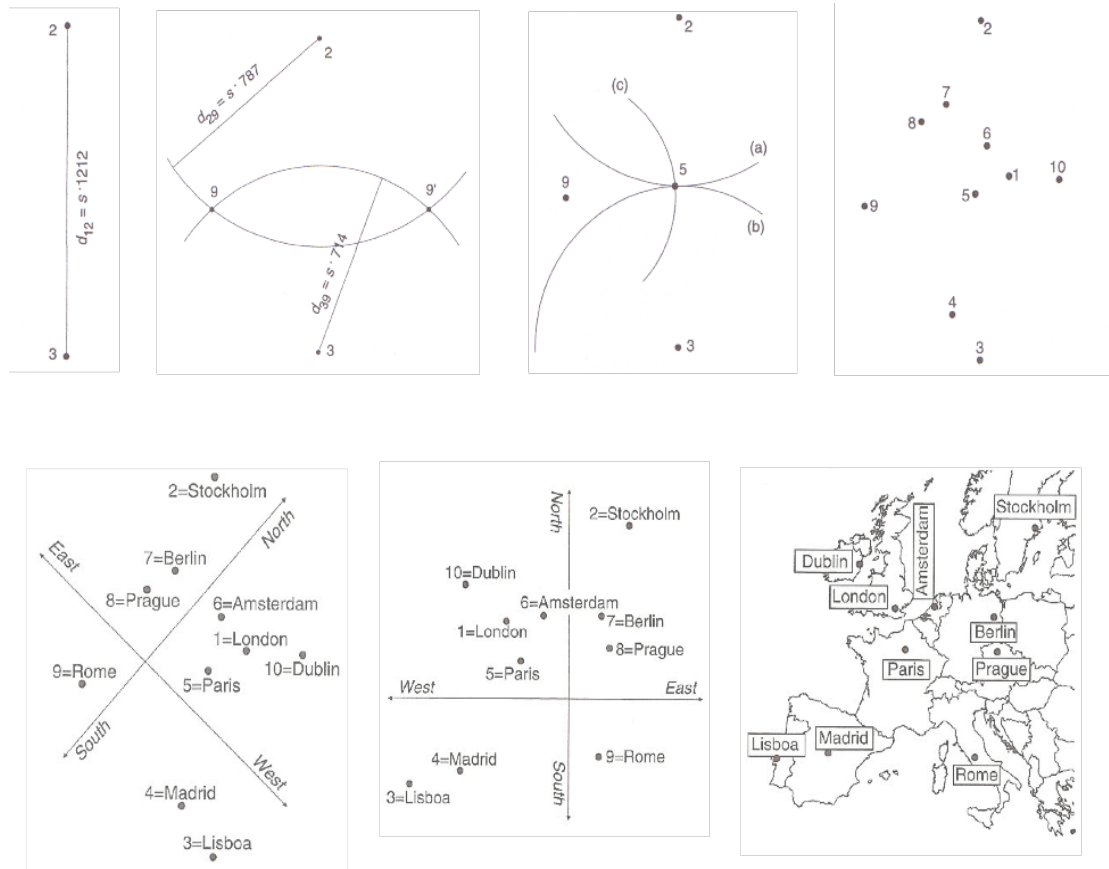
Current understanding is that non-analytical pattern recognition (NAPR) is the predominate mechanism by which expert dermatologists recognise skin lesions and arrive at a clinical diagnosis, but can NAPR be applied by non-experts? In short, do novices have innate pattern recognition abilities enabling them to distinguish differing skin lesions?

This question was investigated with the use of multi-dimension scaling. Multi-Dimensional Scaling (MDS) models were first described in 1952 and allow researchers to explore underlying themes (or dimensions) in order to identify similarities or dissimilarities between the investigated datasets (108). Their use has become widespread within the psychological sciences and informatics. Fundamentally they allow graphical representations of similarity or dissimilarity matrices often using subjective attributes. Although comparison of three attributes is possible using a three dimensional presentation, more commonly two attributes are compared using a two dimensional graph (109). A simple demonstration of their use may be helpful by examining a table of the distances between ten European cities.

	1	2	3	4	5	6	7	8	9	10
1	0	569	667	530	141	140	357	396	570	190
2	569	0	1212	1043	617	446	325	423	787	648
3	667	1212	0	201	596	768	923	882	714	714
4	530	1043	201	0	431	608	740	690	516	622
5	141	617	596	431	0	177	340	337	436	320
6	140	446	768	608	177	0	218	272	519	302
7	357	325	923	740	340	218	0	114	472	514
8	396	423	882	690	337	272	114	0	364	573
9	569	787	714	516	436	519	472	364	0	755
10	190	648	714	622	320	302	514	573	755	0

Table 3.1a: Table of the intercity distances of the ten European cities numbered on Figures 3.1 b-g (109).

By using the distances as a measure of similarity between each of the pairs of points, a graphical display of similarity can be built up. In the resultant plots, which are shown below, the points are arranged so that the distances between any pairs of points have the strongest possible relation to the similarities among the cities which have been paired. In the lower figures the points of the ten cities are rotated according to the compass, then superimposed to give the expected map of Europe.



Figures 3.1b-g: Graphical plots of the intercity distances tabulated in Figure 3.1a. First layer from top left; 3.1b) Cities two & three are located set distance apart. 3.1c-d) Building up the map of the additional eight cities based on similarity/dissimilarity measures. Second layer from bottom left; 3.1e) Names of cities superimposed. 3.1f-g) Points re-orientated to reflect true geographical locations (109).

Similar models are used for subjective similarity matrices; such as in the MDS model below (see Figure 3.1h), where the Scottish Whisky industry have used scores from tastings in which participants graded whisky attributes (light, rich, smoky, delicate) to demonstrate how the peaty Islay malts group together and can be distinguished from the subtler Speyside whiskies. Identifying such associations provides customers with an indication of the whiskies they are most likely to favour.

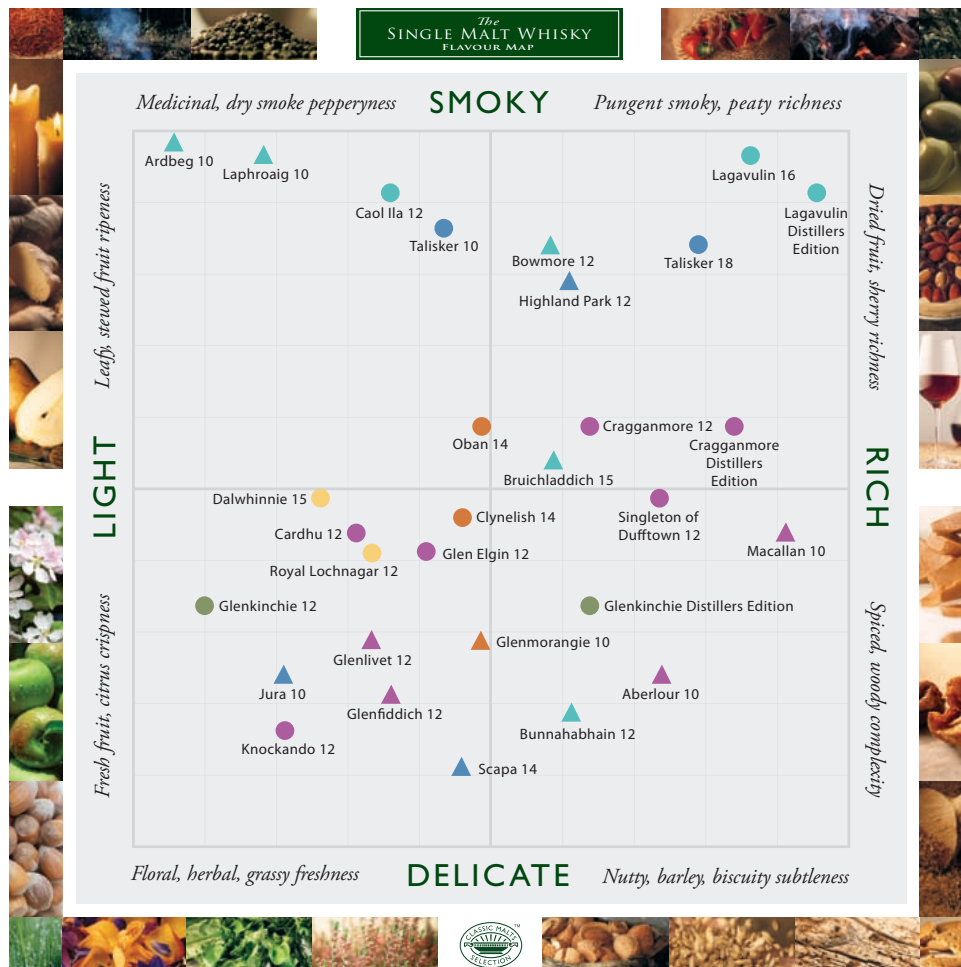


Figure 3.1h: Single malt whisky flavour map (110).

3.2 Aim

In the studies described in this chapter MDS techniques were adapted to identify similarity matrices for novices' perception of skin lesion images.

3.3 Materials & Methods

Subjects

Study 1:

For the first study 20 non-medical lay volunteers were enrolled. The mean age was 36 years old (Range 16-61) and 12 subjects (60%) were female. No volunteers had any personal experience of skin cancer nor had they undergone any tuition in the identification of skin lesions. None of the volunteers were remunerated for their participation.

Study 2:

In the second study a different cohort of subjects was recruited in order to exclude any potential hangover bias from previous lesion exposure. 43 non-medical lay volunteers were enrolled without remuneration. The mean age was 27 years old (18-50) and 17 subjects (40%) were female. Again, none of the subjects had a personal history of skin cancer or skin lesion tuition.

Study 1: MDS pair-wise comparison

The visual similarities of 12 digital images, which had been selected from the Department's image database, were assessed by the novices. The 12 lesions consisted of three images from each of four diagnostic categories of commonly referred lesions; haemangiomas, seborrheic keratoses, melanocytic naevi and basal cell carcinomas (BCCs). The images were selected on the basis of their pathological diagnostic class and the investigators perception of visual similarity. The 12 images used are reproduced below in Figure 3.3a.

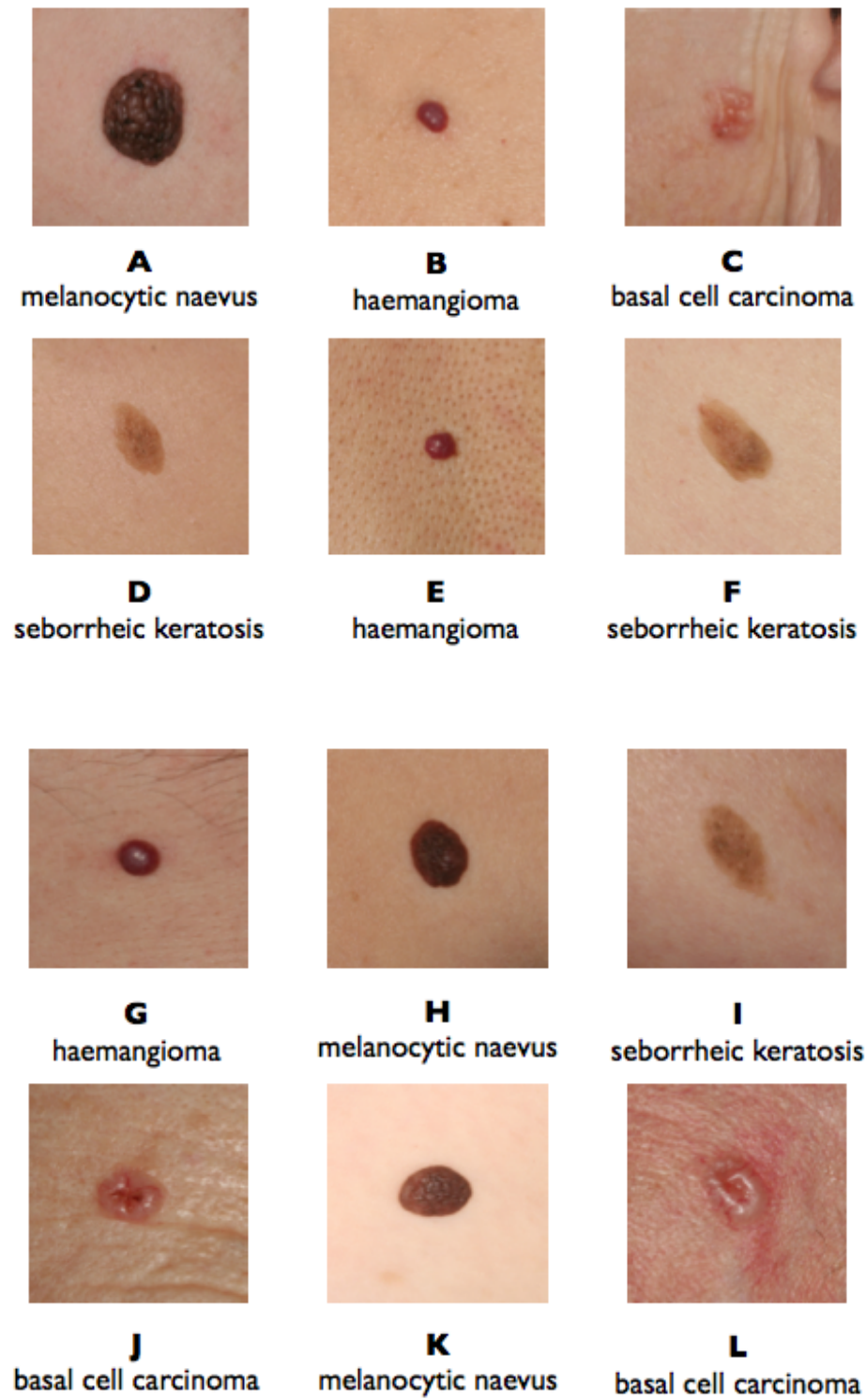


Figure 3.3a: The twelve images used in Study 1 and their diagnostic labels.

The digital images were cropped and printed at identical resolutions (300pixels/inch) and size (55mmx55mm) using a Kodak (KODAK, New York) 1400 Professional Printer onto Kodak Ekatherm Photographic paper. To generate the novices' similarity matrices the subjects were asked to make pair-wise comparisons of each of the 12 lesions with the remaining other 11 lesions, on a 7-point Likert type similarity scale (1=Very Dissimilar through to 7=Very Similar)- see Figure 3.3b below (111).

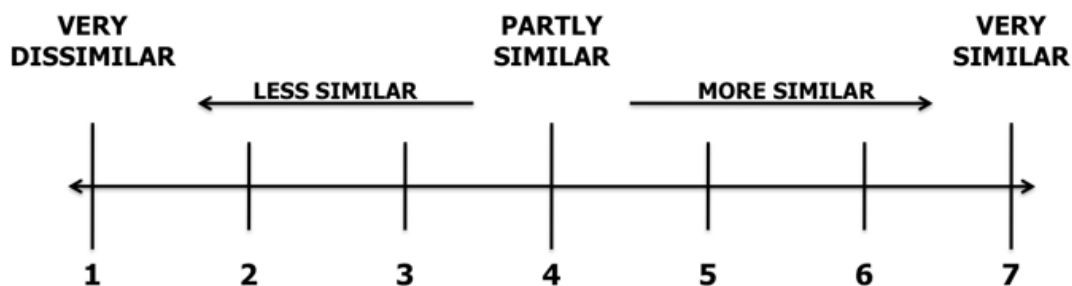


Figure 3.3b: The 7 point Likert like scale that was provided to subjects in Study 1.

This resulted in 66 $((n \times (n-1))/2)$ combinations for each subject. The comparisons were undertaken in the same controlled conditions, placing the two lesions' photographs, equally spaced and in a consistent orientation, on an A4 white testing card, under similar ambient lighting conditions. A reference copy of the scale was visible to the subjects throughout the scoring process. No definition or examples of "similar" were provided. To limit any lead bias or fatigue effect the order in which the 66 pair-wise comparisons were undertaken was randomised for each of the 20 subjects. In addition, the subjects were not allowed to view all twelve lesions together at any point of the experiment.

To test the reproducibility of the subjects' similarity scores, the experiment was repeated at an interval of four weeks with the only modification to the experimental protocol the addition of a Youden pair comparison. A Youden pair operates as an internal control; it involves a specific pair-wise comparison, in which the two stimuli compared are similar but not identical. In this experiment, the same image of a melanocytic naevus was randomly chosen; Lesion A, then orientated at 180° to create

an additional image - Lesion M – for the Youden pair. If the similarity assessment is performed correctly a Likert score indicating a high degree of similarity should be assigned to the Youden pair (112). The images used for the Youden pair used are shown below in Figure 3.3c.

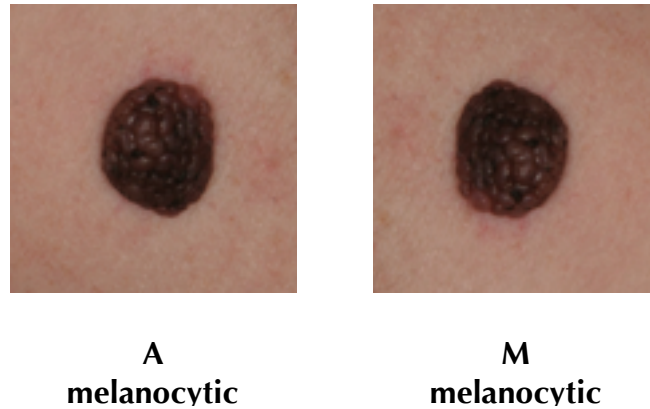


Figure 3.3c: The Youden pair used during Study 1.

After completing the repeated pair-wise analysis each subject was shown all twelve images together for the first time and asked to arrange them into whichever groups they felt to be similar. The subjects were not advised that there were histologically considered to be four groups/diagnoses. The time taken to complete each of the study components was recorded, although the subjects were not made aware of being timed so they did not feel any time pressure constraints.

Study 2: MDS free sorting

This second study was designed to further investigate novices' ability to identify lesional similarity but on both an increased number of lesions and what might be considered a more difficult, visually homogeneous, group of skin lesions. To increase the difficulty of the comparison task and to ascertain whether novices can identify the visual features required to sub-classify dermatological lesions, 30 digital images of lesions from a single diagnostic class were randomly selected from the Dermofit image database. The diagnostic class chosen was that of BCCs, where several visual subclasses (e.g. nodular, infiltrative, ulcerative, superficial) can be discerned by an experienced clinician.

The technique of MDS pair-wise comparisons is inherently inefficient, as it is necessary to assess and score all combinations of lesions to build the similarity matrix. If the same scoring system had been employed in Study two, an individual participant would have been required to make 435 comparisons $((n \times (n-1))/2)$. So, in contrast with the previous experiment, an adaptation of MDS free sorting was applied in a manner similar to the third component of Study one.

The similarity assessments were undertaken remotely over the Internet using custom-built software, designed specifically for this experimental task, to remove any possibility of experimenter interaction bias which might have affected the novices' scoring. The application was developed in Java and deployed on a Tomcat server in conjunction with the Department of Informatics. The software showed the subjects a sequence of ten screens, each screen having two upper target lesions and a set of 24 sample lesions below.



Figure 3.3d: Screenshot to show the purpose-built software used in Study 2. Illustrating some of the randomly selected lesions that the participants were asked to match.

For each of these ten screens the participants were asked to match between zero and six sample images with the target lesions they considered most similar (again no definition of similar was provided). The subjects did this using a drag and drop method as highlighted on the screen grab below.

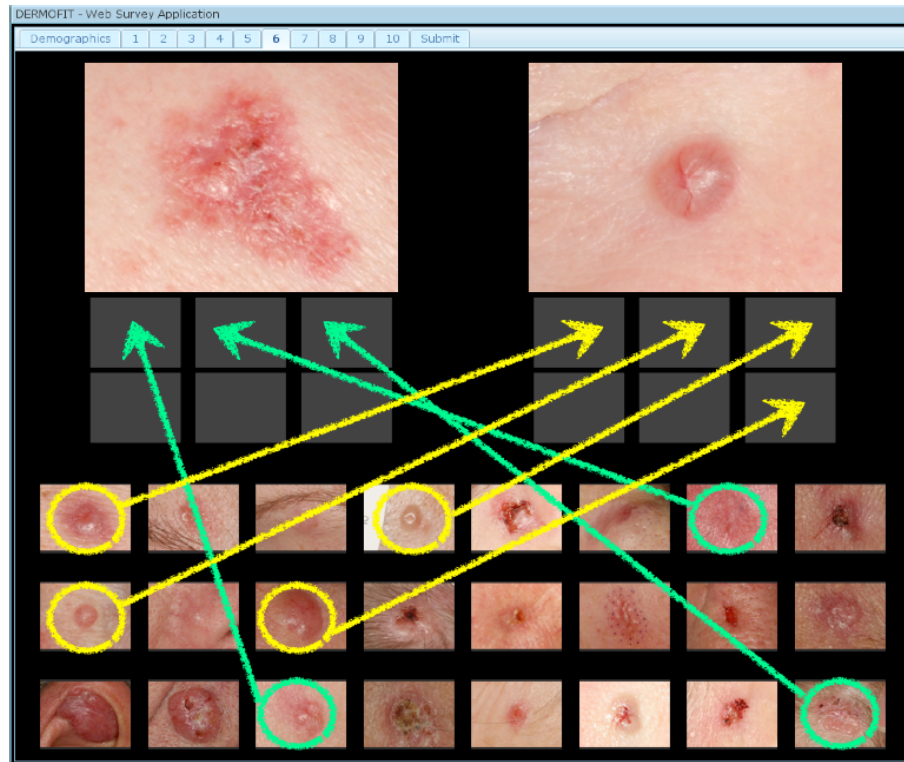


Figure 3.3e: Screenshot to illustrate how the software allowed participants to “drag and drop” images which they considered to match.

The matching data was stored on a specifically designed database created on a PostgreSQL server. Data regarding demographics and the structure of the screens (target and sample images) were saved along with the matches provided by the survey participants. In this free sorting method, each subject’s similarity score for any two images was derived as the percentage of times they had been matched together as similar using the software (i.e. number of occasions matched/number of available opportunities to match).

Statistics

The additional MASS package was installed into R to allow the MDS modelling.

3.4 Results

Study 1: MDS pair-wise comparison

All 2640 Likert similarity scores that the subjects assigned to the pair-wise comparison task are presented below (66 comparisons x 20 subjects x 2 attempts).

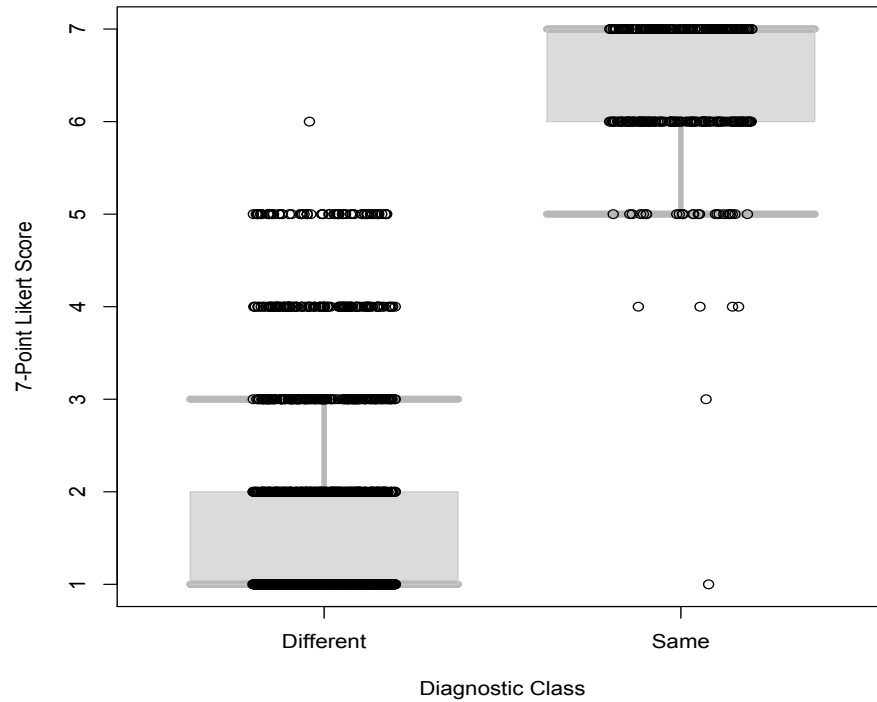


Figure 3.4a: Combined scatter and box plot showing Study 1's 2640 Likert similarity scores.

Pairs of images from the same diagnostic class were attributed a median similarity score of 7 (Very Similar), whereas the pairs of images from different diagnostic classes had a median score of 1 (Very Dissimilar). The mean scores were 6.38 and 1.87 and the modes were 7 and 1 respectively. As can be seen in Figure 3.4a the results are skewed demonstrating how easy the subjects found the task. The difference in similarity scores between intra-class and extra-class comparisons was confirmed statistically significant ($p < 0.0001$, Wilcoxon). The 20 Youden pairs were all given a score of 7 (Very Similar). Intra-person results were consistent between the subjects' two attempts, with a Spearman rank correlation of $Rho = 0.84$ ($p < 0.0001$). Inter-subject results were also consistent between the 66 comparisons with a Kendall's W highly significant ($p < 0.0001$).

The similarity scores were converted to distance matrix and a 2-Dimensional (2D) non-parametric MDS model was derived as a graphical interpretation of the novices' similarity scores.

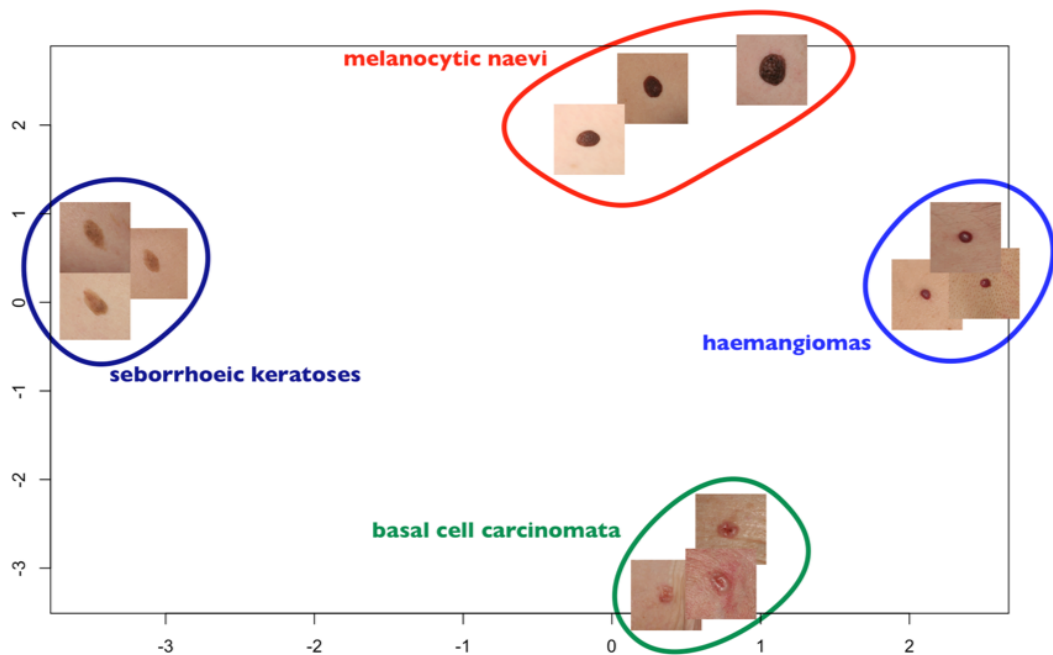


Figure 3.4b: 2D MDS model of novices Likert scores with the images of the individual lesions superimposed on data points.

The model had an excellent fit (as defined as Stress<20%) with a Kruskal's Stress of 7.46%. The MDS model demonstrated clustering of the twelve lesions similar to both the dermatologist's original groupings and to the lesions' pathological diagnoses. In the free-sorting component all 20 subjects correctly split the twelve images into four groups of three lesions corresponding to their pathological diagnosis.

The average time to complete the first 66-pair-wise comparison was 14 minutes (Range 10-31). For the second 67-pair-wise comparisons the average completion time was 11 minutes (Range 8-14). Paired t-test showed these differences to be statistically significant ($p < 0.02$). There was no significant difference between the time taken or the sex of the subjects. All subjects correctly free sorted the twelve images into their correct diagnostic groups. The mean time taken for the twelve lesion free sort was 33 seconds (Range 16-52).

Study 2: MDS free sorting

In total 2395 sample to target matches were performed by the 43 subjects giving a mean target to sample ratio of 1:2.78. The resultant similarity scores for the 30 lesions were converted to a distance matrix and a 2D non-parametric MDS model with a Kruskal's Stress of 19.12% derived.

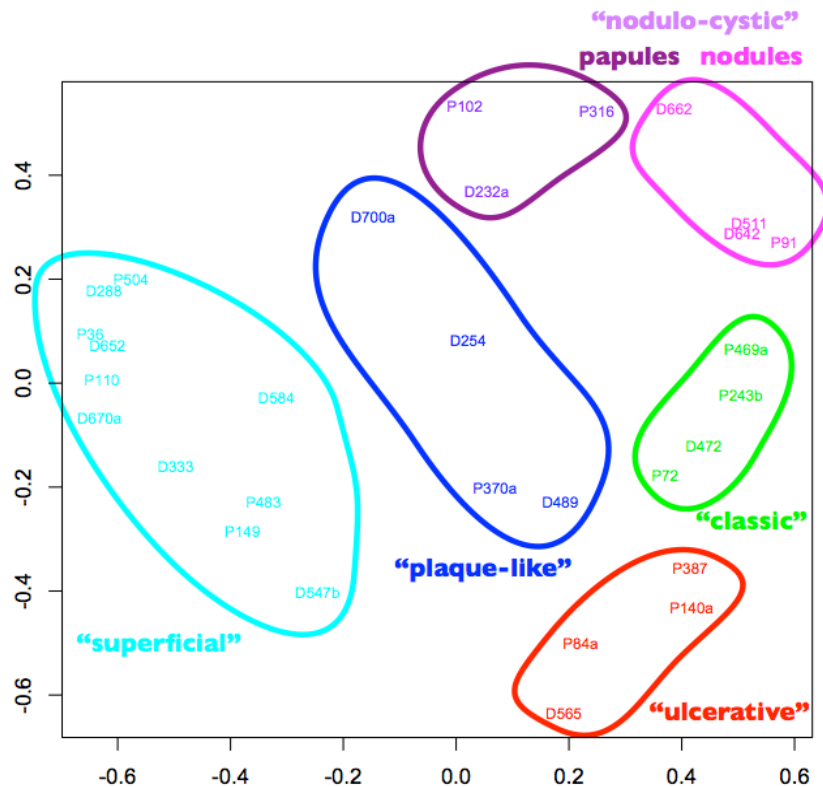


Figure 3.4c: 2D MDS model of novices scoring of similarity for Study 2. The author has superimposed his visual diagnoses of nodulo-cystic, classic, ulcerative, superficial, plaque like BCC sub-types.

All 30 lesions are marked with their identifying file name (the associated images will be presented in Figure 3.4d below) and shaded to indicated the appropriate dermatological classes of the resultant clusters. Although the position of each lesion point on the graph is defined by the similarity data and the interpretation of the MDS model of that data, the colour coded grouping and diagnostic terms appended are those of the author. It can be readily appreciated that novices, without background knowledge or specific education, are able to subgroup lesions of a single diagnostic class into distinct subgroups on the basis solely of their intrinsic perception of visual

similarity. The visual sub-classification of BCCs even by experienced dermatologists is relatively subjective and thus is any interpretation of the resultant MDS model. To present a more objective analysis of the novices' similarity scores the spectral clustering algorithm of Ng was applied to the raw data (113). Hierarchical clustering was then performed in a non-subjective manner, always splitting the bigger cluster into two. The resulting configuration is shown below, providing a remarkable testimony of the ability of non-experts to accurately group similar lesions without any previous training.

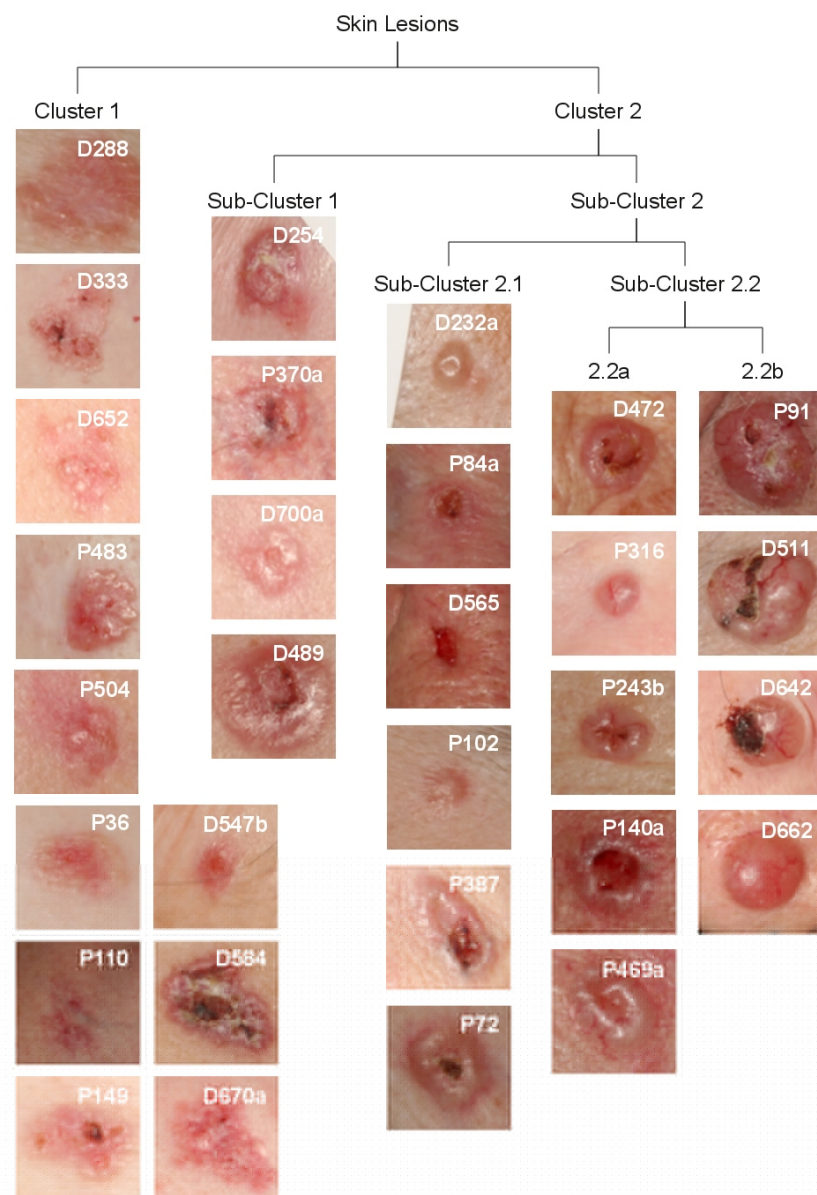


Figure 3.4d: Images of all 30 BCCs used in Study 2 are presented showing the hierarchical clustering resulting from the novices' perception of similarity.

3.5 Discussion

The dual process theory, whereby expert diagnosticians combine intuitive and analytical reasoning to arrive at a diagnosis is widely accepted, albeit the extent to which the analytical system monitors and overrides the intuitive system remains controversial (114). The importance of non-analytical pattern recognition is such that some have suggested that assessment of non-analytical reasoning by methods - such as the puzzle test - may allow selection of those with potentially superior diagnostic ability for medical training (115).

The most common form of intuitive or non-analytical pattern recognition, is overall pattern recognition, which is likely to be the predominate process in dermatological diagnosis. The studies presented in this chapter demonstrate that without any formal tuition lay persons, by employing their innate pattern recognition skills can identify similar skin lesions.

Whilst these studies did not attempt to analyse the cognitive methods or reasoning behind the novices' grouping of lesions as similar, the rapidity with which they did so suggests non-analytical pattern recognition rather than analytical pattern recognition is likely to have predominated. The free sorting of the twelve images, which necessitates that the subjects unconsciously undertake 66 pair-wise comparisons, was accomplished in a mean of just 33 seconds. Furthermore, in study one the subjects completed the second pairwise comparisons in a significantly shorter period of time, providing further support for their analysis having been a learnt non-analytical pattern recognition based response.

Although the first study could be subjected to the criticism that the lesions chosen were visually discrete, the second study demonstrates that even with a more homogenous group of lesions the novices were able to visually sub-classify BCCs in a manner similar to that of experts. The importance of this finding in relation to educational strategies will be subject to further discussions in this thesis.

As described in the methods the generation of the similarity matrices for the MDS model in Study 1 was inefficient due to the necessity of making multiple pair-wise comparisons. Although the efficiency of the process was somewhat improved using the drag-and-drop software in Study 2, MDS investigations remain inherently cumbersome. Whilst MDS models have a utility in small scale studies (such those presented) they are too unwieldy for the large scale ordering of image databases. They are also not suitable for identifying which of the many potential parameters are of the greatest importance to the subjects when making their assessment of similarity.

As a consequence of the inherent inefficiencies of the MDS investigatory model, these studies were unable to determine whether any particular visible parameters exercised a dominant role in the assessments undertaken by the subjects to identify the similarities on which their classification was based. The impression from a review of the images in Figure 3.4d is that no single feature, be it colour, shape, surface contour, contiguity of surface predominates as each cluster shows some features present within the other clusters (albeit some such as ulceration are less prominent in clusters 1.1 and 2.2b). Irrespective of these considerations the overall similarity within the clusters achieved in the studies is striking and accords with the classification of expert diagnosticians. This observation suggests the assessors, in the process of their overall pattern recognition, have taken account of the varying degrees of all visible differences rather than basing their classification upon any single specific feature.

Chapter 4

Novices' analytical pattern recognition ability in distinguishing skin lesions

4.1 Introduction

The work presented in this chapter has been published in Acta Derm Venereol. 2011; 91(2): 125-30. PMID: 21311845. "Novice identification of melanoma: not quite as straightforward as the ABCDs."

Data presented in the previous chapter demonstrated that novices have the intrinsic abilities to group similar lesions, albeit within a narrow experimental framework. The absence of any formal training in those participating in these studies infers similar non-analytical pattern recognition (NAPR) processes as they use for everyday object identification. Most public education strategies intended to enhance non-expert skin lesion identification have concentrated on analytical pattern recognition (APR); with the most common of these APR strategies being the ABCD rule. The studies presented in this chapter were designed to investigate whether novices can utilise APR effectively in discriminating between pigmented skin lesions.

The "ABCD" rule to aid in the diagnosis of early malignant melanoma recently celebrated its 30th Anniversary (116). This mnemonic was introduced in 1985 to aid non-experts' macroscopic diagnosis of pigmented lesions and over the years it has been widely promoted in an attempt to facilitate the earlier detection of melanomas (59). In the thirty years since its inception the use of the ABCD has transitioned from assisting non-expert physicians' diagnosis, through educating the public in intensive clinician led programs, to its present central position in most public education strategies; the criteria being described on the public websites of the American Academy of Dermatology (AAD) (117), British Association of Dermatologists (BAD) (118), Australasian College of Dermatologists (ACD) (119) and European Academy of Dermatology and Venereology (EADV) (120). Although the fundamental four-part

ABCD mnemonic has received widespread adoption, surprisingly there has been little validation of its utility as a general public education strategy.



THE ABCDEASY WAY TO CHECK MOLES

There are three main types of skin cancer, and each looks different. This guide tells you a few of the signs that might indicate a 'melanoma' - the deadliest form of skin cancer. Tell your doctor about any changes to your moles or skin, even if they are not the same as those described here.

- **ASYMMETRY**
The two halves of the mole may differ in their shape and not match
- **BORDER**
The outside edges of the mole or area may be blurred, and sometimes show notches or look 'ragged'
- **COLOUR**
This may be uneven and patchy. You may see different shades of black, brown and pink
- **DIAMETER**
Most, but not all, melanomas are at least 6mm in diameter. If any mole gets bigger or changes, tell your doctor
- **EXPERT**
Look out for change, and if in doubt, check it out! If your doctor is worried about your skin, make sure you see a Consultant Dermatologist, an expert in diagnosing a skin cancer. Your doctor can refer you for free through the NHS.

IF IN DOUBT, CHECK IT OUT!
www.bad.org.uk | 020 7391 6084
Registered Charity No: 258474



Figure 4.1a: Public sun awareness campaign approved by BAD (118).

Given that the majority of melanomas are first identified by lay individuals, whose tardiness is also responsible for the largest proportion of any delay in their diagnosis, reliable and accurate information to assist the public in early detection is essential (121-124). Recently the importance of accurate patient information has been further highlighted as, thus far, the numerous strategies initiated to improve screening and the early detection of melanoma have not resulted in a substantial reduction in the proportion of tumours diagnosed with prognostically unfavourable thickness (125).

There is now mounting evidence to question the use of the ABCD rule as a general public education strategy (126,127).

The majority of studies that are cited as providing evidence for the mnemonics' adoption have either involved clinician-performed assessments (128-132), or intensive lay education (133,134), and, as examined below, there are methodological limitations in extrapolating the findings from such studies to general public health promotion, in particular with relation to the roles of experience and exposure to prior examples.

The role of experience

In the identification of skin lesions, as previously described, experts including dermatologists appear to predominately employ NAPR, derived through exposure to prior examples. Such overall pattern recognition techniques cannot be unlearned and thus have a carry-over effect on any attempts to assess the application of analytical rules (135-139). Furthermore, study designs in which experts are asked to explain their diagnoses inherently over-emphasize algorithmic reasoning by promoting intentional rather than incidental analytical processing (64). As already discussed it has even been suggested that experienced clinicians make a diagnosis intuitively then alter their analytical assessment to accord with their preconceptions about the relationship between these features (such as the ABCD) and the diagnosis (such as melanoma) (64,140). Certainly, the only prospective study of dermatologists' recognition patterns, confirmed that whilst analytical pattern recognition (the ABCD rules) could be used to predict malignancy, it was not actually how dermatologists arrived at the correct diagnosis; instead the experts seemed to use overall or differential pattern recognition ("Ugly duckling sign" (60)) (56).

The role of prior examples

The exact number of prior examples that are required to significantly alter analytical assessments is unknown, but these carry-over effects do not only apply to seasoned clinicians, as novices have been shown to exhibit this bias with prior exposure to only

a few examples (136,137,141). Whilst we do not fully understand how novices intuitively visually assesses skin lesions, unless it significantly differs from the rest of human visual perception it is unlikely to be by the application of analytically based reasoning. If, as appears likely, overall-pattern recognition plays a role, be that even a small role, the biasing effect of prior examples must be controlled when assessing the effectiveness of the analytical ABCD criteria. In the few studies where intensive ABCD education has been demonstrated to have a beneficial effect on lay individuals, the effect of prior examples and overall pattern recognition was not controlled (133,134). It is therefore unclear how much of the positive effect of the ABCD rule can be attributed to the novices' ability to discriminate using the true analytical ABCD criteria, rather than their ability to use overall-pattern recognition 'learnt' from the example lesions employed to demonstrate how to apply the analytical criteria during their education. Thus far the only randomized control trial testing the ABCD in lay hands showed that it decreased diagnostic accuracy and was not as effective as a pattern recognition education strategy (126).

In this context, to establish that any screening or diagnostic tool has value ideally three criteria should be satisfied: Inter-observer variability should be minimal; variation within a diagnostic class should be small; and the inter-diagnosis differences should be significant.

4.2 Aim

In the study presented in this chapter the three subjective analytical properties (Asymmetry, Border Irregularity, Colour difference) i.e. the ABCs of the ABCD rule are examined against the following criteria:

- 1) inter-observer variability
- 2) variation within a diagnostic class
- 3) inter-diagnosis differences

4.3 Materials & Methods

Subjects

An open email request containing the uniform resource locator (URL) link to the study was distributed to MSc students of the University of Edinburgh's School of Informatics, inviting them to personally undertake the study and forward on to non-medical acquaintances who also might be willing to participate. 33 lay subjects agreed to participate without remuneration. Sex distribution was split with 21 males and 12 females (64% male). Mean age was 34 years old (Range 17-62). None of the subjects had a personal history of skin cancer.

Study Image selection

40 digital images of pigmented skin lesions were randomly selected from 878 relevant images in the Dermofit image database. The 40 images selected for this experiment were stratified so that there were 10 images from each of the following 4 diagnostic classes: benign naevi, dysplastic naevi (defined as lesions with histological atypia), melanomas and seborrheic keratoses. The lesions were cropped from the original digital photograph, each to the same resolution (600x450 pixels) and displayed in a 1:1 ratio (approximately equivalent to seven times magnification from the original 50cm capturing distance) in the custom built experimental software.

Software design & implementation

A purpose-built program was created to allow the three subjective criteria of the ABCD rule (Asymmetry, Border Irregularity and Colour Uniformity) to be tested remotely over the Internet, to limit any possible investigator-related interaction bias. Diameter was felt not to be suitable for assessment as the images were magnified and thus would have required the placement of a relative 6mm marking scale next to each lesion before asking each participant to comment which was longer; the lesion or the measuring scale, which would not have been instructive.

The program was written in JavaScript and PHP (Personal Home Page) in conjunction with the Department of Informatics, and after entering their demographics the participants were given a set of instructions explaining how to use the ABC criteria and the software. The instructions were based on the verbal descriptors of the ABC criteria taken from the website of the AAD (117). After confirming that they had read and understood the instructions the subjects assessed each of the 40 images in turn for Asymmetry, Border Irregularity and Colour Uniformity on a 10-point visual analogue scale (VAS). A screen shot of the software can be seen below. At any stage the subjects could click on a “Help” button to redisplay the instructions and verbal descriptors of the ABC.

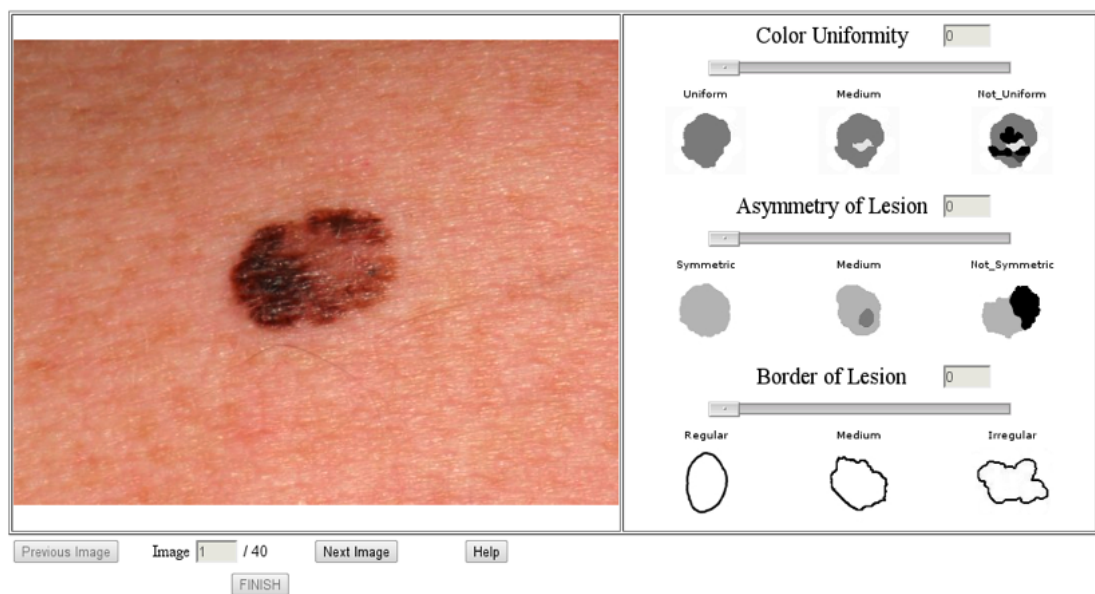


Figure 4.3a: Screen shot of purpose built software used in the non-experts’ assessment of the 40 lesions for the ABC’s of the ABCD criteria.

In addition to the text descriptions of the three properties to be assessed, in order to further reduce inter-rater differences the visual anchor points shown in Figure 4.3a were also integrated into the software at the mid-point and ends of the rating scale. The high-end anchor points were taken from the ABCDE patient guidelines on the

SkinCancerNet “Melanoma: What it Looks Like” webpage produced in conjunction with the AAD - see Figure 4.3b for the screenshot from this webpage (142).

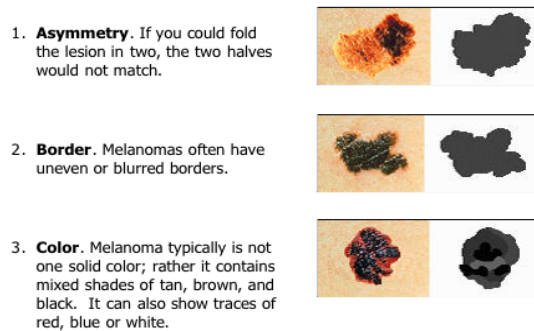


Figure 4.3b: Screenshot from the SkinCancerNet “Melanoma: What it looks like” webpage produced in conjunction with the AAD. This provided the high-end anchor points for the non-experts’ scoring (142).

As the study was designed to assess the lay public’s ability to discriminate analytically these three properties and not their ability to use their innate non-analytical reasoning to match or distinguish from real-life examples, only the caricatured diagrams from the SkinCancerNet website were used as the high-end anchor points, rather than the accompanying clinical pictures of melanomas. The mid and low-end anchor points were computer generated to complete the VAS (See Figure 4.3c).

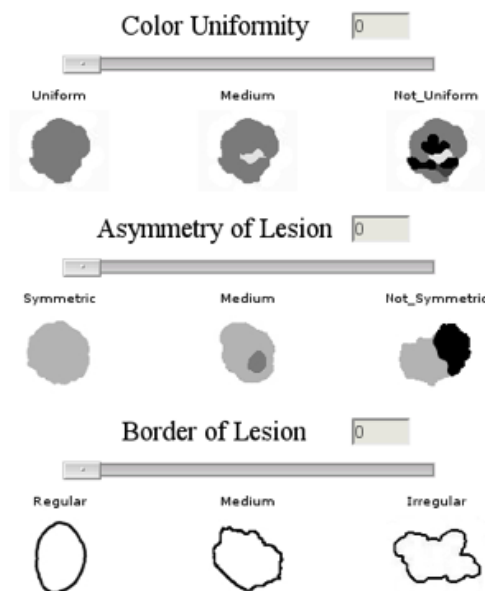


Figure 4.3c: The complete VAS (visual analogue scale) used by the non-experts’ for assessing the ABC’s of the ABCDs.

To minimize any lead bias or fatigue effect, the software was programmed to randomize the order in which the subjects undertook the 40-lesion assessment, to ensure it varied for each participant.

4.4 Results

The full results of all 3960 analytical VAS scores attributed in the study (33 subjects x 40 lesions x 3 'ABC' properties) are graphically displayed in Figures 4.4 a-c, with each property presented in a separate plot. At first glance these plots may seem complicated but they have the virtue that every data-point from the experiment is presented and the variability in scoring can be instinctively comprehended.

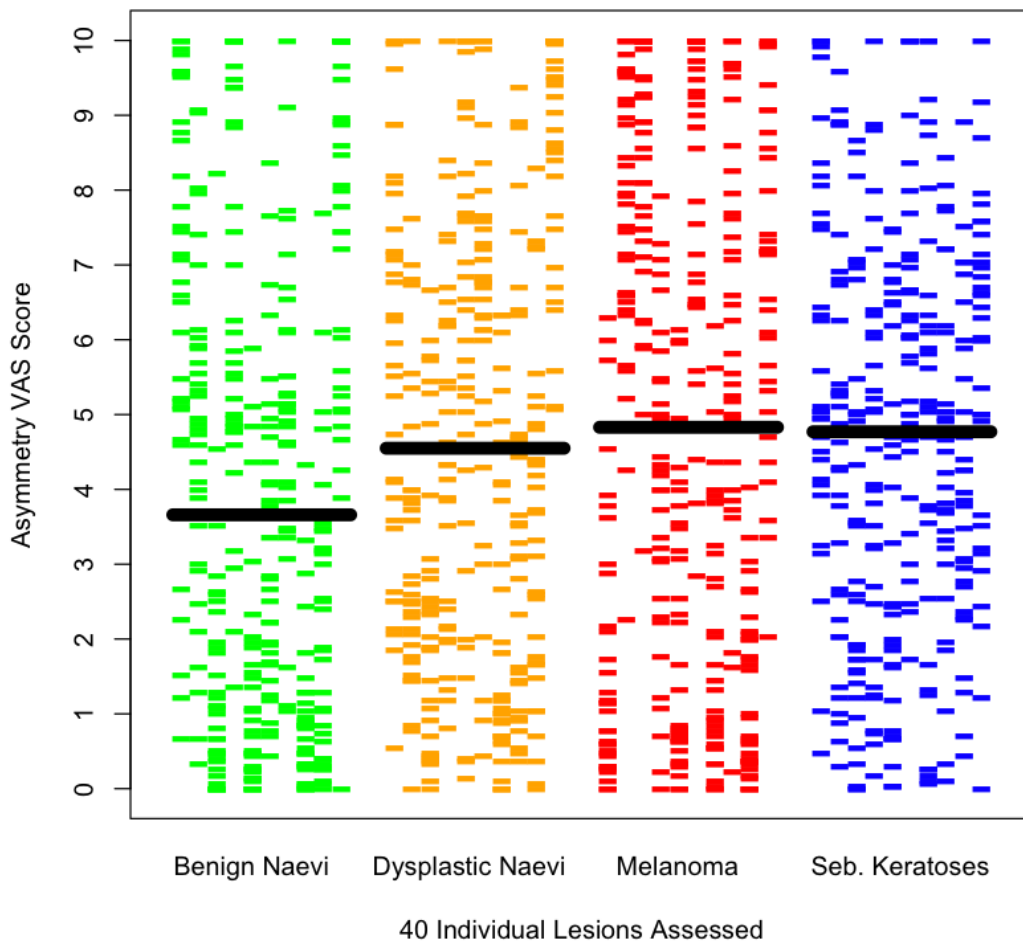


Figure 4.4a: Plot of all 1320 Asymmetry VAS scores attributed by the 33 subjects to the 40 lesions.

In explanation: each horizontal coloured bar represents an individual subject's score for a specific lesion, with each of the 40 lesions displayed in individual columns across the X-axis. These 40 lesions' columns are grouped into different colours according to their diagnostic classes (Green= the 10 Benign Naevi, Orange= the 10 Dysplastic Naevi, Red= the 10 Melanomas, Blue= the 10 Seborrheic Keratoses). The overall median score for each diagnostic class is shown by the large horizontal black bar, straddling each of the 4 coloured series of 10 columns.

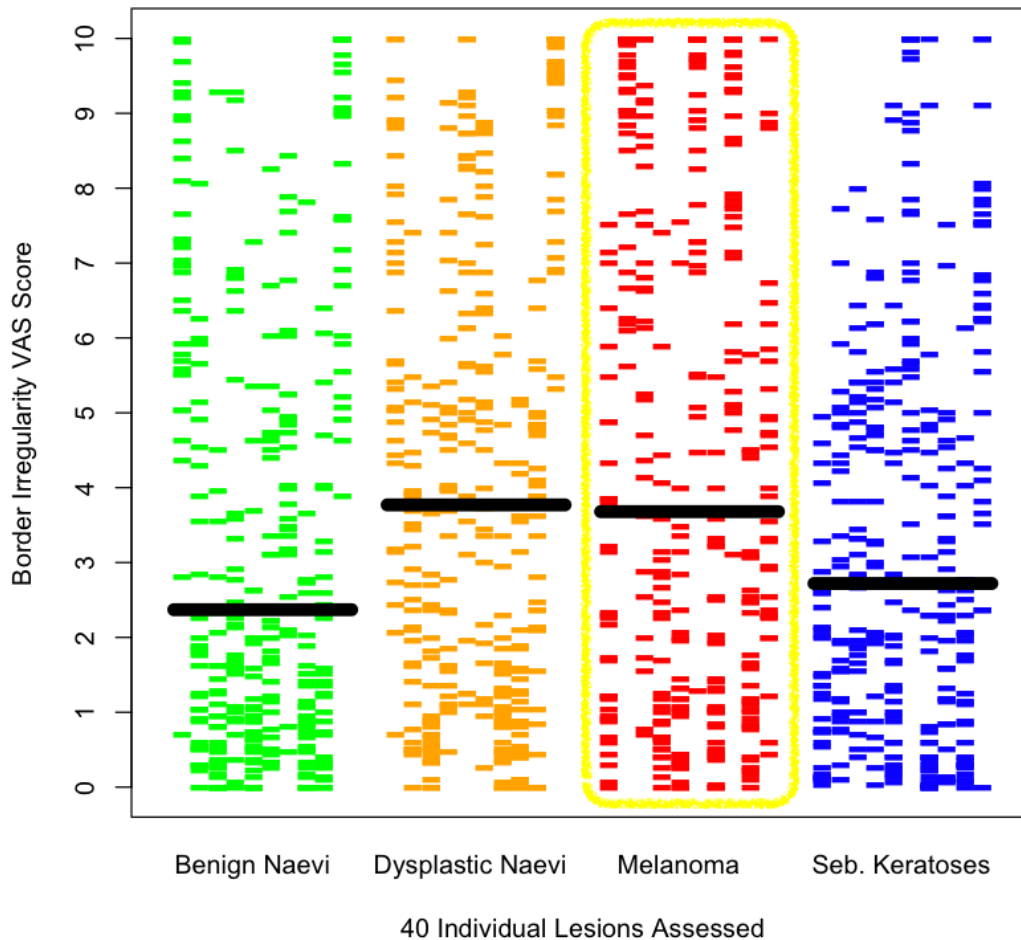


Figure 4.4b: Plot of all 1320 Border Irregularity VAS scores attributed by the 33 subjects to the 40 lesions. The highlighted section in yellow is reproduced in greater detail in Figure 4.4e.

The inter-person variability in assessing each of the 3 ABC properties for any single lesion is represented by a single column's vertical spread across the Y-axis. Within a

specific diagnostic class the variability in scores is demonstrated by the differences in vertical spread between the 10 uniform coloured columns.

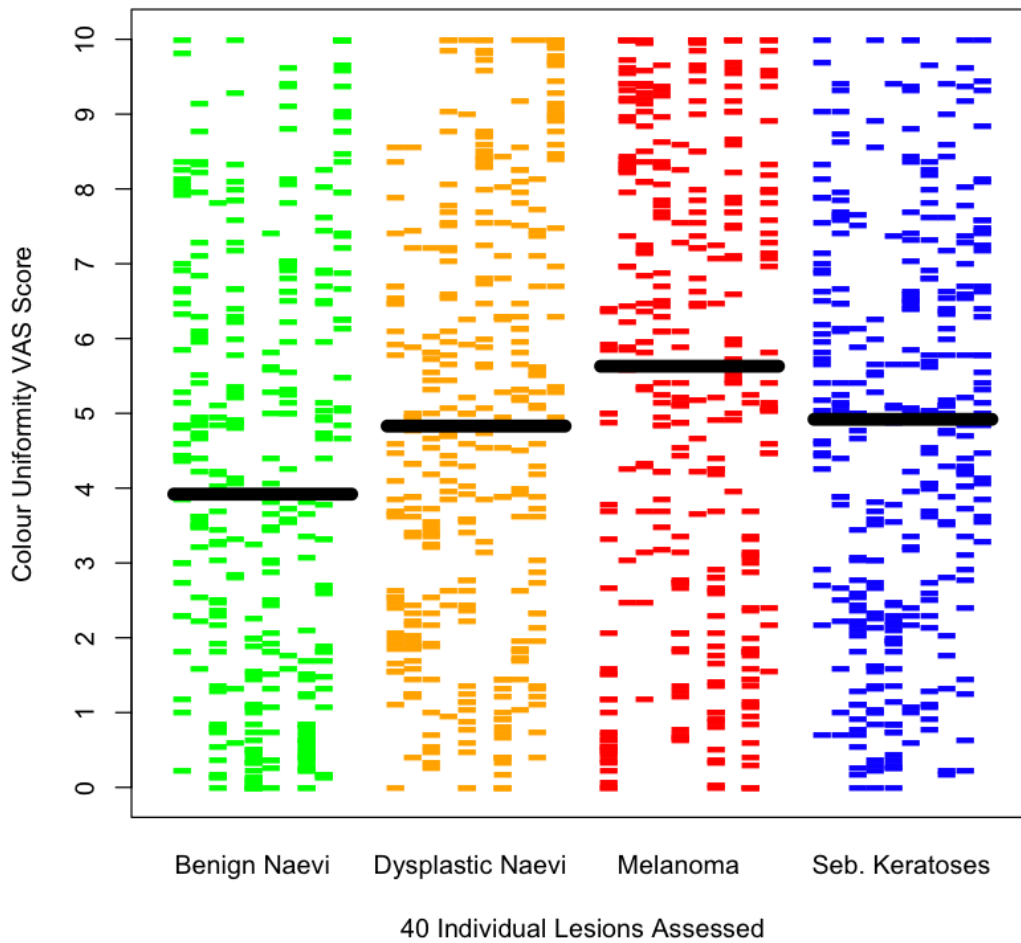


Figure 4.4c: Plot of all 1320 Colour Uniformity VAS scores attributed by the 33 subjects to the 40 lesions.

The variability between the 4 diagnostic classes is appreciated by the differences in the overall distributions between the 4 coloured groups and further enforced by the differences in their median scores indicated by the horizontal black bars. The results are further summarised in Table 4.4d below.

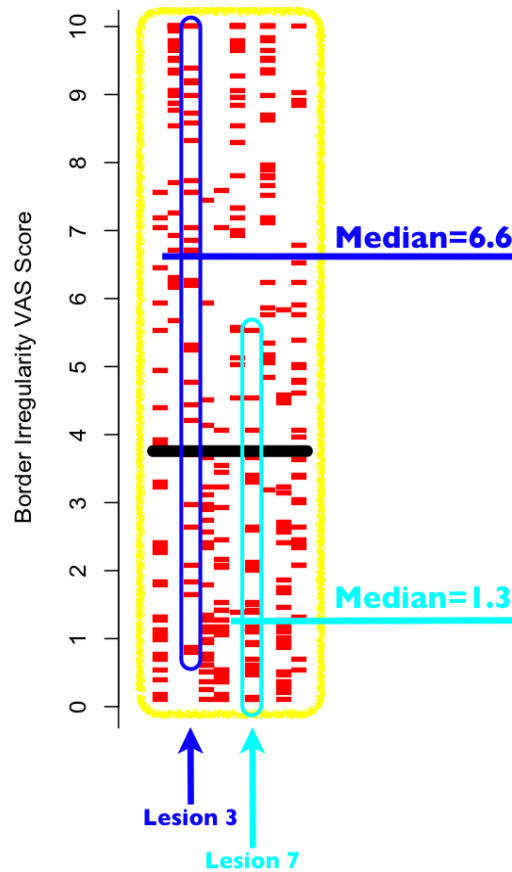
Lesion Class	Benign Naevi	Dysplastic Naevi	Melanomas	Seborrheic Keratoses
Asymmetry VAS Scores (0= Symmetric, 10=Asymmetric)				
Median	3.66	4.55	4.83	4.77
IQR	4.96	4.88	5.94	4.07
90 th Percentile	8.93	8.97	9.52	8.09
Border Irregularity VAS Scores (0= Regular, 10=Irregular)				
Median	2.37	3.77	3.68	2.72
IQR	4.61	5.05	6.46	4.08
90 th Percentile	8.27	8.93	9.78	7.52
Colour Uniformity VAS Scores (0= Single uniform colour, 10= Multiple or non-uniform colour distribution)				
Median	3.92	4.83	5.63	4.92
IQR	5.00	4.88	5.59	4.23
90 th Percentile	8.26	9	9.59	8.34

Table 4.4d: Table of additional data showing the spread of the subjects' responses. The median, inter quartile range (IQR) and 90th percentile of the scores are broken down by ABC feature and the lesions' diagnostic class.

Whilst it is possible to appreciate the small, albeit significant (Kruskal Wallis= $P < 0.0001$), difference between the 4 diagnostic groups' scores perhaps what is far more striking is the substantial spread in the scores attributed to the same lesion by the 33 subjects and the further variation in scoring between the 10 lesions within the same diagnostic class for all 3 of the subjective ABC properties. Analysis of additional data demonstrates that a similar substantial variation exists within the 10 scores that each individual attributed to the lesions within the same diagnostic class.

To better enable appreciation of the inter-person and intra-class variation Figure 4.4e presents an enlarged view of the highlighted section of Figure 4.4b. In Figure 4.4e it

can be observed that lesion 3 (blue arrow/highlights) which had the largest inter-person variation (IQR= 4.86) of the 10 lesions within the melanoma class had a range of 'border irregularity' scores attributed by the 33 subjects from 0.7 to 10 with a median score of 6.6, and lesion 7 (cyan arrow/highlights), which had the least inter-person variance (IQR=1.93), had a range from 0 to 5.5 with a median score of 1.3. This can be further appreciated in Table 4.4f which shows the median VAS border scores and IQRs for all of the 10 lesions displayed in Figure 4.4.e.



Lesion	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
Median VAS score	2.3	9.3	6.6	1.2	1.4	8.9	1.3	7.9	1.1	3.9
IQR	3.4	1.4	4.9	2.3	1.7	2.9	1.9	3.3	1.7	3.3

Figure 4.4e & Table 4.4f: Highlighted section of Figure 4.4b to demonstrate the inter-person and intra-class variation of the VAS border scores attributed by subjects. Table 4.4f provides the individual lesion data with both the median VAS score and IQRs per lesion.

4.5 Discussion

The principal motivation for undertaking the study described in this chapter was to use the ABC rule as proxy for APR in instructing non-experts in the identification of pigmented lesions and assessing their subsequent ability to use such analytical methods to distinguish skin lesions.

The original rationale and justification for the ABCD approach had been to aid the primary target group of non-specialist physicians in the diagnosis of melanoma, but the rule has subsequently been extended to the lay public despite there being little supporting evidence to justify the transfer. This development generates at least two concerns:

First, an increasing body of evidence is accumulating that experts are not necessarily able to explicitly identify the parameters on which their own expertise is based in a manner that is transferable (64,140).

Second, as previous testing of ABCD had not controlled for prior exposure (128-134), it is possible that the claimed utility of the ABCD may have reflected prior experience and knowledge rather than the implementation and use of the criteria itself (135-139).

A further relevant consideration is that whereas experts may be able to use the criteria on appropriate subclasses of lesions (i.e. melanocytic nevi and melanomas), distinguishing primary melanocytic lesions from their many mimics (such as seborrhoeic keratoses) is likely to require considerable more expertise than the mastery of generic rule based clinical features (95,143,144).

The experimental approach taken was an attempt to delineate the characteristics of the ABCD rules on a test series of lesions. As outlined above, for the ABCD system to be capable of guiding diagnosis, it would have to be demonstrated statistically that the different diagnostic groups scored differently, that variance between persons for the same lesion and within diagnostic groups were small. Within the limits of the test situation and the randomly chosen images it is self-evident that these criteria were not

met. Individual subjects judged the same lesion very differently, and although the medians of different diagnostic groups differed, the overlap was considerable. Examining the graphs of the three properties, it is difficult to imagine being able to choose any criteria based on ABCs that would usefully discriminate suspicious from banal lesions.

Some limitations were present in the experimental model. The subjects were not chosen at random as all were highly educated, computer literate, and likely to possess above average abstract and analytical reasoning abilities. It is doubtful, however, that any bias inherent in such a selection invalidates the results.

Second, not only were primary melanocytic lesions included but also mimics such as seborrhoeic keratoses. The justification for latter being that non-experts and many non-specialist physicians are unable to reliably distinguish between these classes of lesions and, in practice, the study recognises that the ABCD criteria will be applied (incorrectly) to non-melanocytic pigmented lesions (95).

Third, as the subjects undertook the experimental task remotely over the internet it was not possible to assess the 'effort' they applied in determining their scoring. However, because the order in which each subject assessed the 40 lesions was randomized any fatigue effect would have been minimized. Indeed, close inspection of the data demonstrates that whilst there is substantial overlap in scoring between (and within) the diagnostic groups, the subjects' scoring was not random; individual lesions having (to varying degrees) distinct scoring patterns.

Although it is possible had the subjects undergone intense education in the use of the ABCD approach the results may have been different; as the promulgation of the ABCD criteria via web sites and patient leaflets provides little opportunity for intense education, such speculation is irrelevant to assessing the impact of current practice on early diagnosis of malignant lesions.

As highlighted above previous studies assessing the benefits of the ABCD approach have been methodologically compromised because of failure to control for exposure

to relevant clinical images during the teaching phase. Unlike these previous assessments, the novices in this study were denied exposure to examples of clinical images during their ABCD training. The failure of the presented study to demonstrate any diagnostic advantages using the ABCD may reflect that exposure to sample images during the course of training persons in the use of the ABCD in previous studies led to the development of pattern recognition skills from even the minimal exposure to such images. Any improvement in performance identified in those studies may therefore have reflected the acquisition of NAPR skills over APR. To challenge this interpretation requires a degree of experimental control lacking in previous studies. Furthermore this conclusion is supported by the findings of the only large scale RCT of lay education where the ABCD training approach was found not to be as effective as image based education (126).

A number of modifications to the basic ABCD mnemonic have been suggested over the years to “improve” its functionality, and it is now commonplace to use the ABCDE criteria (145-154). Across the literature there is inconstancy in the meaning of the ‘E’, with evolving, enlarging, elevated, erythema, expert all being proposed. Given the evidence presented in this chapter that untrained novices cannot effectively employ the analytical ABC criteria, should not the public education message now be simplified to heighten concerns over a lesion that is Evolving (i.e. Changing). Such a simplification has previously been suggested by Weinstock (152,153) and has independently been found to be the most important predictor of melanoma in features observed by patients (127). Of note, since publishing the work presented in this Chapter others have started to question the utility of the role of ABCDE as public education strategy (155).

Chapter 5

The education of medical students in the diagnosis of skin lesions

5.1 Introduction

The work presented in this chapter has been published in BMC Medical Education 2012; 12:27. PMID: 22569037. "Dermatology undergraduate skin cancer training: A disconnect between recommendations, clinical exposure and competence."

The importance of early diagnosis of skin malignancies has already been discussed and irrespective of what direction future skin lesion screening may take, at the present time UK general practitioners are primarily responsible for the initial assessment of skin lesions acting as the gatekeepers to dermatological cancer services (98,99). The only formal clinical dermatology teaching for most medical practitioners is as an undergraduate. Despite cutaneous disease having been shown to be the most common reason for primary care consultations (at 24% of physicians' new patient workload) (156), medical school training in dermatology is relatively limited comprising only approximately 1% of undergraduate teaching time (103,104). This chapter investigates the current state of undergraduate skin lesion training at the University of Edinburgh.

Whilst historically each UK University has set its own curricula, taking account of the recommendations from the General Medical Council (GMC), in 2006 the British Association of Dermatologists (BAD) issued the document "Dermatology in the Undergraduate Medical Curriculum", outlining the dermatological skills expected of students upon graduation (157). These guidelines were based on the results of a modified Delphi study that had been conducted with a multi-disciplinary panel of 66 individuals involved in delivering undergraduate education (158). Under the "skin cancer" section the guidelines state that it is "very important" that graduates should be able to recognise the three main skin tumours (basal cell carcinoma (BCC), squamous cell carcinoma (SCC) and melanoma) and that it is "fairly important" that they can

recognise another thirteen common skin lesions, but as it must always be “very important” that students are able to distinguish the rarer malignant tumours from the more frequent benign lesions there is, by implication, a degree of incompatibility in these aims. A 2009 audit of UK medical schools’ curricula concluded that these “skin cancer” learning outcomes were present in the syllabuses of most universities (159). Although the inclusion of a topic in a curriculum cannot be taken to assume that for any particular student those competencies are actually achieved.

5.2 Aim

Given the paucity of research work in this area, the study presented in this chapter describes the collection of evidence of the competencies of students in the University of Edinburgh across this clinical domain and attempts to relate them to clinical exposure and diagnostic confidence. The students’ exposure to skin cancers was determined to ascertain whether the prescribed competencies are being achieved.

5.3 Materials & Methods

Subjects

Study 1:

All 77 students undertaking their dermatology attachment between July and September 2009 were enrolled. Fifty students (65%) were female. Over the ten-week study period, 5 cohorts of between 14-17 students attended.

Study 2:

In study two, the total clinical exposure for each of the 50 students who attended their undergraduate dermatology attachment from July to August 2010 was assessed. Over a six-week period three cohorts of between 16-17 students were monitored.

Study 3:

In study three all students completing their dermatology attachment between July and September 2010 (n=50) and again between March and May 2011 (n=61) were enrolled.

Study 1: Student competence at skin tumour diagnosis

The students' diagnostic accuracy was examined using a digital slideshow, with the students writing diagnoses on a custom designed answer sheet. An answer was "generously" accepted as correct irrespective of spelling mistakes, incomplete terminology, abbreviations and the use of lay terms. Assessment was undertaken prior to seeing any patients on the first morning of the attachment (Day 1 test) and again on the final afternoon (Day 10 test). Note that the students are not formally examined at the end of each two-week dermatology block, but later undertake a joint written exam with other disciplines.

Five separate test batches of images were assembled so that both the Day 1 and Day 10 tests differed for each cohort of students. Each test batch consisted of 25 digital images randomly selected in a stratified manner to ensure that each test contained an identical spectrum of lesional diagnoses. The order of the 25 lesions was randomised and included one or more examples of each of the sixteen lesions considered important in the BAD guidelines. The test images were randomly selected from 687 suitable digital photographs in the Dermofit image database.

The digital slideshows were conducted in an identical manner using iPhoto on a 15" Apple Macbook Pro (APPLE, California) connected to an Epson H285B high definition projector (EPSON, Japan), under similar ambient lighting conditions in the Department's seminar room. The 25 lesions were all presented at the same scale for 20 seconds each. Students were not given feedback after the Day 1 test, but after completing the Day 10 test attended an addition tutorial during which feedback was provided.

Twelve months later, in August 2010, 30 of the original students were randomly selected, contacted by email and asked to return to the department. The students were not informed of the reason for their selection but they were advised that they would be helping with a research project and would receive a £10 honorarium for their participation. Nineteen students (63%) were able and willing to attend for retesting (12 Month test). The batches of test images used for the 12 Month test were coordinated so

that each student re-attending had not been exposed previously to the same digital photographs. The slideshow was conducted in an identical manner as described above for the Day 1 and Day 10 tests.

Study 2: Clinical exposure to skin lesions during the attachment

Every clinical visual encounter that the students were exposed to was recorded on a purpose-designed checklist. This checklist contained the same 16 skin lesions considered important in the BAD guidelines and was completed by each student for their “one-on-one” teaching and skin surgery sessions, whereas for the group demonstration clinics it was completed by a dedicated clinical observer. If a diagnosis was not on the checklist additional space was available to record text for “other” diagnoses. If more than one lesion was seen on the same patient each lesion counted as a separate clinical encounter. In contrast, if a patient was being followed-up or being used to demonstrate a “dermatological history”, unless there was also a visual example of a skin lesion, these patients did not count as a clinical encounter under the terms of reference of the present study. Medical staff involved in the teaching clinics were not pre-warned that the monitoring process was being undertaken.

Students were asked to complete an anonymous questionnaire on the final afternoon of their attachment (Day 10). The questionnaire asked them to rate their confidence on 7-point Likert like scales across the three key skin cancer learning outcomes; their ability to diagnose melanomas, their ability to diagnose basal cell carcinomas and their ability to diagnose squamous cell carcinomas. See Figure 5.3a below for Likert scale used.

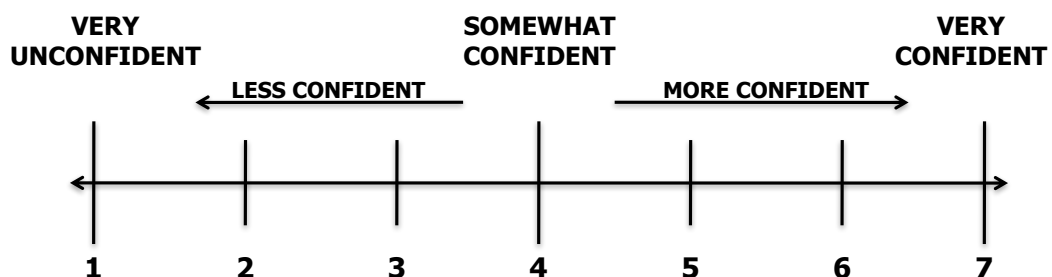


Figure 5.3a: The 7-point Likert like scale that was provided to subjects in Study 2 to self-rate their confidence to correctly diagnose the three main skin cancers.

Study 3: Learning exposure outside the clinic

To gain a measure of other non-patient visual dermatology exposure, all students completing their dermatology attachment between July and September 2010 (n=50) and again between March and May 2011 (n=61) were asked to complete an anonymous questionnaire designed to identify the additional resources used to supplement their formal teaching. In total 106 (95%) students completed the questionnaire. The purpose-designed questionnaire had tick boxes covering the most popular textbooks and Internet sites with space to add supplementary free-text if the resource used was not listed. In addition to assessing the resources used the questionnaire also asked the students to estimate how long on average they spent consulting these resources.

5.4 Results

Study 1: Student competence at skin tumour diagnosis

Of the 77 students enrolled, 74 and 70 completed the Day 1 and Day 10 tests respectively and their scores for the Day 1 and Day 10 tests are shown in Figure 5.4a.

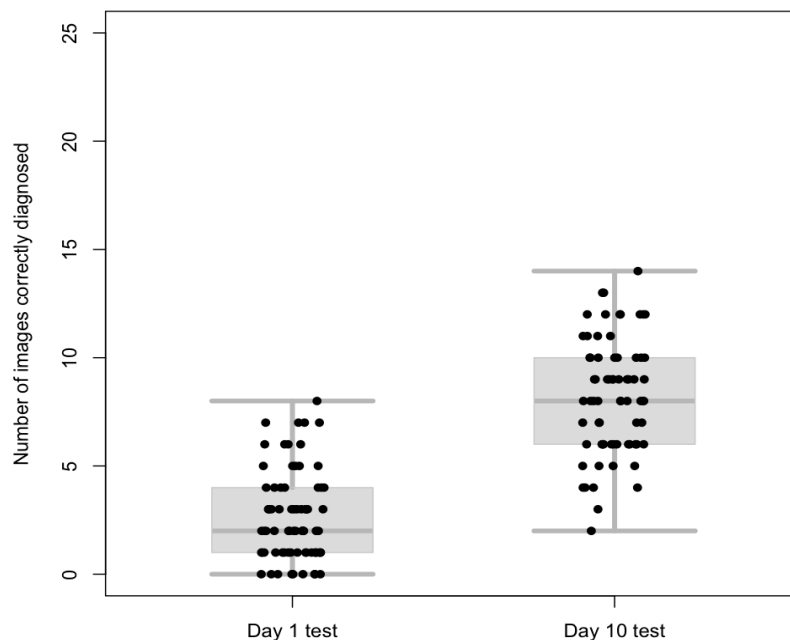


Figure 5.4a: Combined scatter and box plot showing the students' scores for the Day 1 and Day 10 tests from Study 1.

In the Day 1 test, out of the 25 test images, the median number of images correctly identified was 2 (Range 0-8), with an overall diagnostic accuracy of 11% (195/1850). In the Day 10 test the median number of images correctly diagnosed was 8 (Range 2-14), with an overall accuracy of 33% (575/1750).

Sixty-seven (87%) students completed both the Day 1 and Day 10 tests and the difference between these students' results on Day 1 and Day 10 was significant ($p < 0.0001$, Paired Wilcoxon). The overall Effect size ($\mu_1 - \mu_2 / \sigma$) of the dermatology attachment on students' diagnostic accuracy was +2.72 (160). No differences in diagnostic accuracy were observed between the sexes, the cohorts of students or the individual 25 image test batches.

The scores for the 30 students that were randomly selected for recall after 12 months are shown in below in Figure 5.4b.

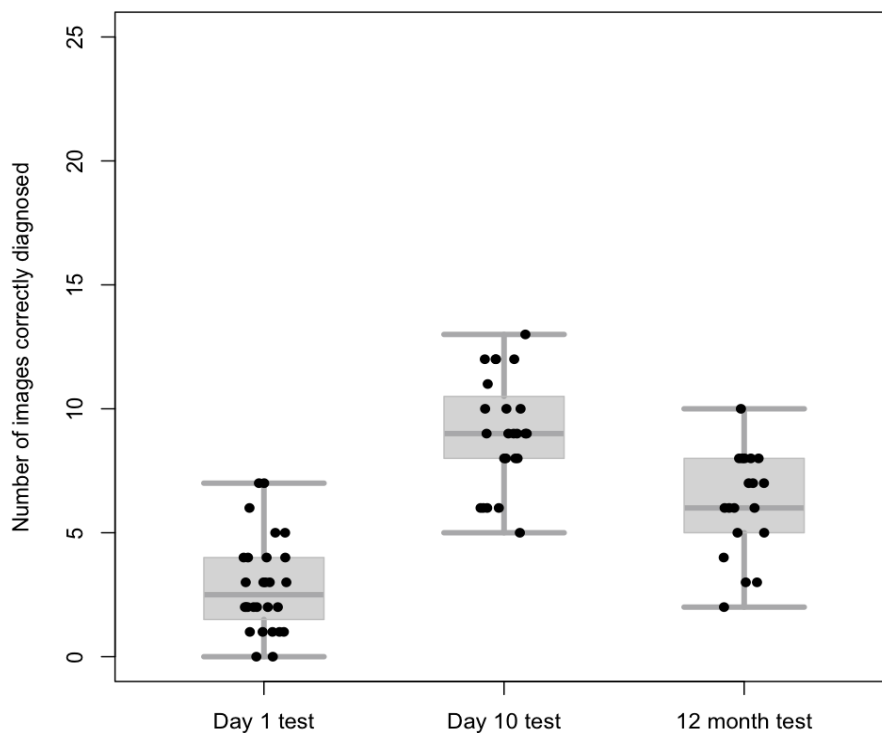


Figure 5.4b: Combined scatter and box plot showing the students' scores (recalled after 1 year) for the Day 1, Day 10 and 12 month tests from Study 1.

In the Day 1 test these students achieved a median score of 2.5 (Range 0-7) with an accuracy of 11% (80/700) and in the Day 10 test a median score of 9 (Range 5-13) with an accuracy of 36% (217/600). After a year their median score was 6 (Range 2-10) with an overall accuracy of 24% (117/475). A Linear mixed-effects model confirmed that there were significant differences between the Day 1 and Day 10 scores, the Day 10 and 12 Month scores and the Day 1 and 12 Month scores ($p < 0.001$ for each). For this recalled subgroup the initial effect size was +3.1 dropping after a year to +1.59. Again, in the test at 12 months, no differences were observed in the diagnostic accuracy between the sexes, the cohorts of students or the individual test batches.

Inspection of the data indicates that the 19 students who attended for the 12 Month test were representative of the original 77 students enrolled in terms of their sex distribution and original scores on both the Day 1 and Day 10 tests. Similarly, the 11 students that were randomly selected but unable to attend for re-testing appeared similar in their original diagnostic scores to the 19 students that were able to attend.

Study 2: Clinical exposure to skin lesions during the attachment

The 50 students saw a median of 23 lesions (7-36) during their dermatology attachments. The three cohorts of students differed significantly in the number of lesions witnessed with medians of 23, 18 and 31 (Kruskal-Wallis $P < 0.001$). (See Figure 5.4c below).

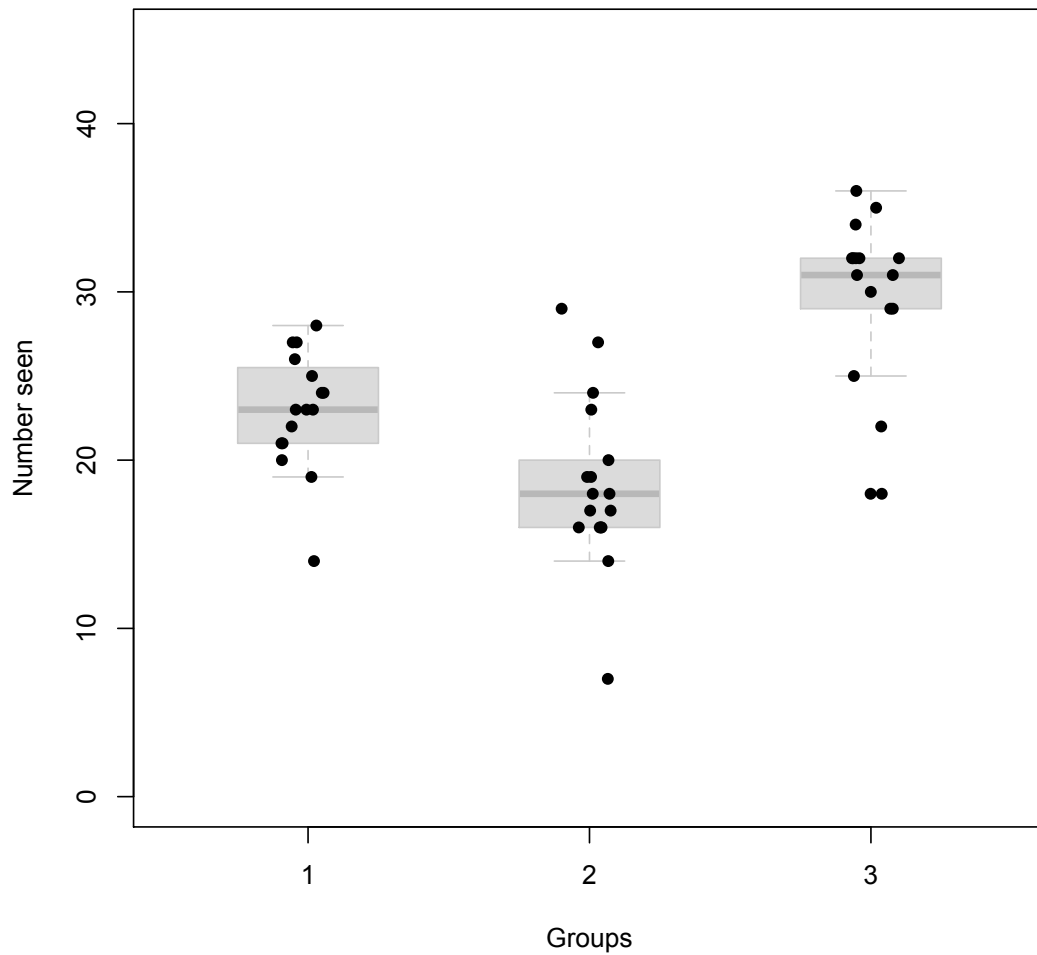


Figure 5.4c: Combined scatter and box plot showing the number of skin lesions students witnessed across the three cohorts of students examined in Study 2.

For the sixteen benchmark diagnoses highlighted in the BAD guidelines the median number of these seen by each student was 8 (Range 5-11) which did not vary by student cohort. None of the students witnessed an example of each of the sixteen lesions listed in the BAD guidelines. The overall observation rate, as defined as the number of positive visual encounters that the 50 students witnessed across the sixteen recommended lesions, was only 53% (426/800). A full breakdown of the results are in Table 5.4d below. The percentage of students witnessing 1, >3 and >5 examples is given for each of the sixteen lesions and demonstrates that there was not only a lack of breadth but also of depth to the students' exposure.

UK Recommended Lesions	Student cohort 1 (n=16) median exposure	Student cohort 2 (n=17) median exposure	Student cohort 3 (n=17) median exposure	Overall student median exposure n=50 (Range)	Percentage of students observing 1 or more example of each lesion	Percentage of students observing >3 examples	Percentage of students observing >5 examples
Individual Lesions "Very Important"							
BCCs	6	2	6	5 (0-11)	98%	70%	42%
SCCs	0	2	1	1 (0-4)	76%	2%	0%
Melanomas	1	0	0	0 (0-2)	38%	0%	0%
Individual Lesions "Important"							
Viral Warts	0	0	0	0 (0-1)	28%	0%	0%
Epidermoid cysts	0	0	0	0 (0-1)	10%	0%	0%
Melanocytic Naevi	1	1	5	2 (0-7)	72%	30%	14%
Seborrhoeic Keratoses	6	4	8	6 (1-10)	100%	78%	56%
Solar Keratoses	3	4	5	4 (1-7)	100%	60%	12%
Bowen's disease	1	1	2	1 (0-4)	96%	6%	0%
Dermatofibromas	1	1	1	1 (0-3)	80%	0%	0%
Keratoacanthomas	1	2	0	1 (0-2)	64%	0%	0%
Lipomas	0	0	0	0 (0-1)	6%	0%	0%
Pyogenic granulomas	1	0	1	1 (0-2)	54%	0%	0%
Mycosis fungoides	0	1	0	0 (0-2)	28%	0%	0%
Paget's disease	0	0	0	0 (0-1)	2%	0%	0%
Cutaneous metastases	0	0	0	0 (0-0)	0%	0%	0%
All 16 lesions							
Total number of lesional diagnoses seen	9/16	8/16	8/16	8 (5-11)	426/800 (53%) Overall Observation Rate	123/800 (15%)	62/800 (8%)
Total number of skin lesions seen	23	18	31	23 (7-36)	NA		

Table 5.4d: Table detailing the students' exposure to skin lesions in Study 2.

The most common lesions witnessed were solar and seborrhoeic keratoses with all the students exposed to at least one example of each. The students exposure to the three main skin cancers is plotted below in Figure 5.4e.

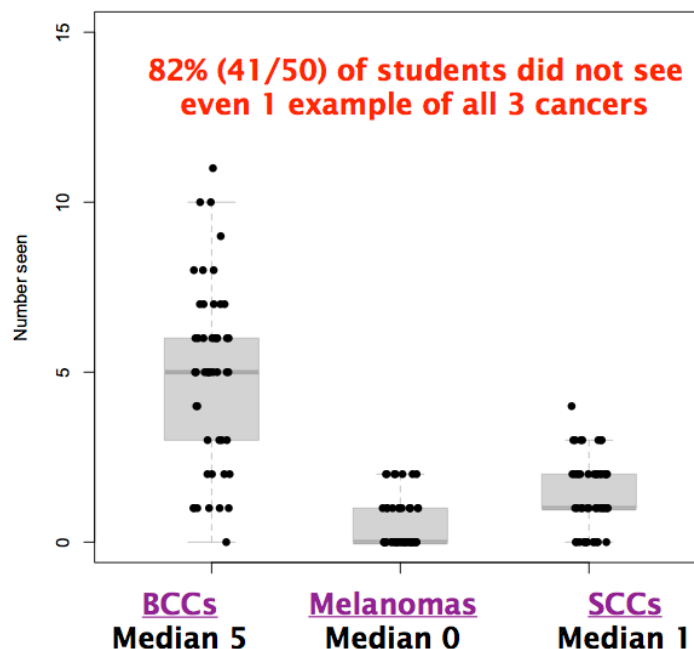


Figure 5.4e: Combined scatter and box plot detailing the students exposure to the three "very important" skin cancer lesions.

The median number of BCCs seen was 5 (Range 0-11) with 98% (n=49) of students witnessing at least one example, for SCCs the median witnessed was 1 (Range 0-4) and 76% (n=38) of students saw an example. For melanomas only 38% (n=19) of students saw an example with the median witnessed being 0 (Range 0-2). The overwhelming majority of students 82% (n=41) did not see an example of each of the three major skin cancers (BCC, SCC, melanoma) and only a single student (2%) witnessed two examples of each.

Forty-four students (88%) completed the end of attachment questionnaire. The scores for this student self-assessment questionnaire are presented in the table below. At the end of their attachment only 34%, 14% and 27% of students described themselves as not confident at identifying melanomas, BCCs and SCCs respectively. Despite the different levels of exposure between the three cohorts of students there was no significant difference between their confidence scores.

Questions: How confident are you in your ability to diagnose...	Mean Likert Score (Range)	Median Likert Score	Percentage of students "unconfident" (score <4)
...melanomas?	4.2 (2-6)	4	34% (n=15)
...squamous cell carcinomas?	4.3 (2-7)	4	27% (n=12)
...basal cell carcinomas?	4.8 (2-7)	5	14% (n=6)

Table 5.4f: Table showing students' Likert confidence scores in diagnosing the three skin main skin tumours. For a copy of the Likert scale used see Figure 5.3a.

Study 3: Learning exposure outside the clinic

Ninety-two percent (97/106) of students used additional textbooks of whom the majority, 58% (56/97), used more than one book. It is worth noting that the University of Edinburgh has an extensive dermatological library adjacent to the Dermatology Department to which the students have full access to during their attachment. 65% (69/106) of students used online resources but only the minority, 43% (30/69), of these students used more than one site. The total number of students using each resource is given along with the relative percentage of students for that type resource (either

internet or textbook) are presented in the table below. The percentages do not add up to 100% as some of the students used multiple resources. The most popular resource was the New Zealand Dermatological Society's Website (161). Students reported spending a median extra 1-2 hours/day studying dermatology during their attachment.

Internet resources (n=69)	
Website	No. of Students (%age)
DermNet NZ (161)	59 (85%)
University of Edinburgh (162)	25 (36%)
Wikipedia (163)	12 (17%)
BAD (164)	5 (7%)
E-medicine (165)	2 (3%)
PCDS (166)	1
Patient.co.uk (167)	1
Textbook resources (n=97)	
Textbook	No. of Students (%age)
Davidson's Principles and Practice of Medicine (168)	57 (59%)
Clinical Dermatology (169)	53 (54%)
Dermatology: an Illustrated Colour Text (170)	31 (32%)
Oxford Handbook of Clinical Specialties (171)	14 (14%)
ABC of Dermatology (172)	9 (9%)
Crash Course (173)	5 (5%)
Physical Signs in Dermatology (174)	4 (4%)
Fitzpatrick's Dermatology in General Medicine (175)	3 (3%)
Differential Diagnosis in Dermatology (176)	2
Rook's Textbook of Dermatology (177)	1
Lecture Notes: Dermatology (178)	1

Table 5.4g: Table of additional resources used by the students in Study 3.

5.5 Discussion

The first study demonstrates that students' diagnostic accuracy for the sixteen "important" skin lesions was only 33% after completing their dermatology attachment and that it dropped to 24% a year later. Such absolute measures of diagnostic competence are obviously deficient and can be potentially misleading because they are so dependent on the difficulty of the test examples, furthermore it is unlikely that such measures are a ratio scale. However, in the absence of normative data elsewhere in the literature, they provide a basic benchmark of diagnostic competence.

A possible reason for the students' poor accuracy was identified in the second study, where it was found that students' exposure to the sixteen important skin lesions is highly variable but universally limited. For the reasons discussed below some limitation in clinical exposure is to be expected, but the limited extent of students' clinical exposure was surprising. Although there is no data available on this issue it is likely that the clinical exposure gained by medical undergraduates at the University of Edinburgh is equal to, if not better, than that at many other UK centres, as the duration of the dermatology attachment is greater than the UK average and there is only a single department in the city with one of the largest throughput of patients in the UK (103-105).

Due to the structure of the group demonstration clinics (which allows multiple students to see a single patient selected from the other concurrent departmental clinics by reason of their good clinical signs) the overall melanoma witnessing rate of 38% was predominately achieved because of a single case -if this one patient had not presented and agreed to be examined at the time of a demonstration clinic the overall melanoma observation rate would have dropped to 18%.

The students' low exposure demonstrates the major difficulty encountered in teaching students to identify skin cancers and their mimics which is the lack of consistent clinical examples. Unlike other specialties where pertinent clinical signs are often longstanding and can be demonstrated repeatedly to subsequent groups of students, suspected skin malignancies are excised as a matter of priority and benign lesions are

promptly discharged back to the community. As a result, the traditional model of face-to-face patient teaching is unlikely to provide a viable solution, as the throughput of relevant new patients is difficult to coordinate and unequally distributed. The variability of the latter is illustrated in this study, where notwithstanding that 38% of this student cohort (unusually) witnessed a melanoma, 76% of students witnessed a SCC and 98% witnessed a BCC, the distribution of these lesions within the teaching clinics was such that 82% did not see an example of all three tumours. Furthermore, this method of teaching cannot be practicably extended to all groups requiring tumour education.

It might be argued that “skin cancer” education is not performed exclusively during students’ dermatology attachments as additional teaching may occur during their other clinical attachments. On balance this seems unlikely as the students’ scores for the diagnosis of common skin lesions were lower one year after their dermatology training, notwithstanding that assessment followed their two-month attachment in primary care.

Although it was statistically not possible to link the individual student’s diagnostic confidence to their personal accuracy data, there would seem to be a mismatch between their objective diagnostic ability and their self-confidence. This may not be surprising, as although subjective measures are in widespread use for assessing teaching and learning, there is increasing acknowledgment that students’ (and indeed other health professionals’) self-assessment of their own abilities can often be erroneous (179,180).

A possible reason for the students’ overconfidence is that the lesions encountered during their teaching attachment or the clinical illustrations in their additional resources were not suitably representative of the broader spectrum of presentations that were randomly selected for our test slide shows from the extensive Dermofit library (which are arguably more representative of those encountered in real-life). Neither the lesions that the students witness in the group demonstration clinics nor those that they are exposed to in their additional study materials are unselected, with the lesions in both textbooks and in the demonstration clinics often chosen precisely

because they are “classic” examples. This matter will be further considered in a Chapter 8. Irrespective of the cause, the mismatch in confidence and competence apparent at the completion of the students’ only formal dermatology training has implications for the role of non-experts gatekeepers to cutaneous cancer services (98,99).

Given the potential importance of undergraduate dermatology education it is surprising that there has been so little investigation of UK medical school dermatology teaching (103,104,159,181-186). Whilst additional studies have subjectively shown that both medical students and primary care physicians feel dermatology training could be improved (187-189), there are only a few US studies that have objectively assessed students’ diagnostic acumen after undergraduate dermatology electives (190-192). The results of these studies are, however, unlikely to be transferable to UK undergraduate students, and are not directly comparable with this study as the consistency in the difficulty of the questions was uncontrolled.

Whilst the US studies demonstrated a higher level of improvement in diagnostic competence (71-82% accuracy), potentially there may be several reasons for this:

First, in these studies there could be an element of selection bias; the attachments were not compulsory with the students enrolling having volunteered to undertake dermatology electives, many presumably with the intention of pursuing a dermatology career.

Second, the studies used the same images for both the initial and final assessments, which are likely to have artificially raised the final scores. More importantly the test contained not just lesions but also dermatoses and the images used for testing were not randomly selected, instead, the images chosen were often described as “classical” examples.

Third, the answers were in multiple choice format rather than free text, which does not correspond to everyday practice.

Finally, these optional electives were of longer duration than in the UK, being full-time for 2-4 weeks.

Of note, since publishing the work presented in this chapter, a recent meta-analysis of interventions into skin cancer education examined 47 outcomes (across 37 studies) and concluded on the resultant Forrest plot that the work outlined here was the 4th most effective intervention reported in the literature (193).

Chapter 6

The learning curves for the diagnosis of skin lesions

6.1 Introduction

The work presented in this chapter details investigations designed to explore methods to improve the identification of skin lesions. It has been established that students seek to supplement their dermatological teaching with online resources presumably as the number of images accessible online far exceeds that available in the clinic or in textbooks. The most recent formal investigations into dermatological online images have shown a high accuracy at 98% (194), however, online content may vary in quality with many online images being of poor resolution often with a questionable diagnosis as in the screen shot examples below – see Figures 6.1a&b (195,196).

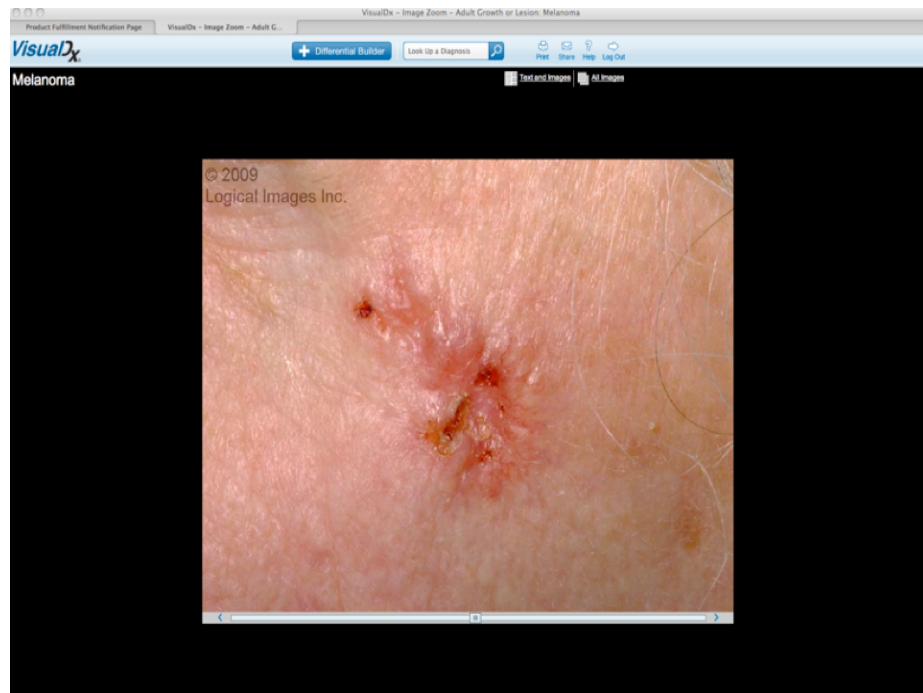


Figure 6.1a: Screenshot showing the first “melanoma” example from VisualDx website- an NHS funded image databank (195).

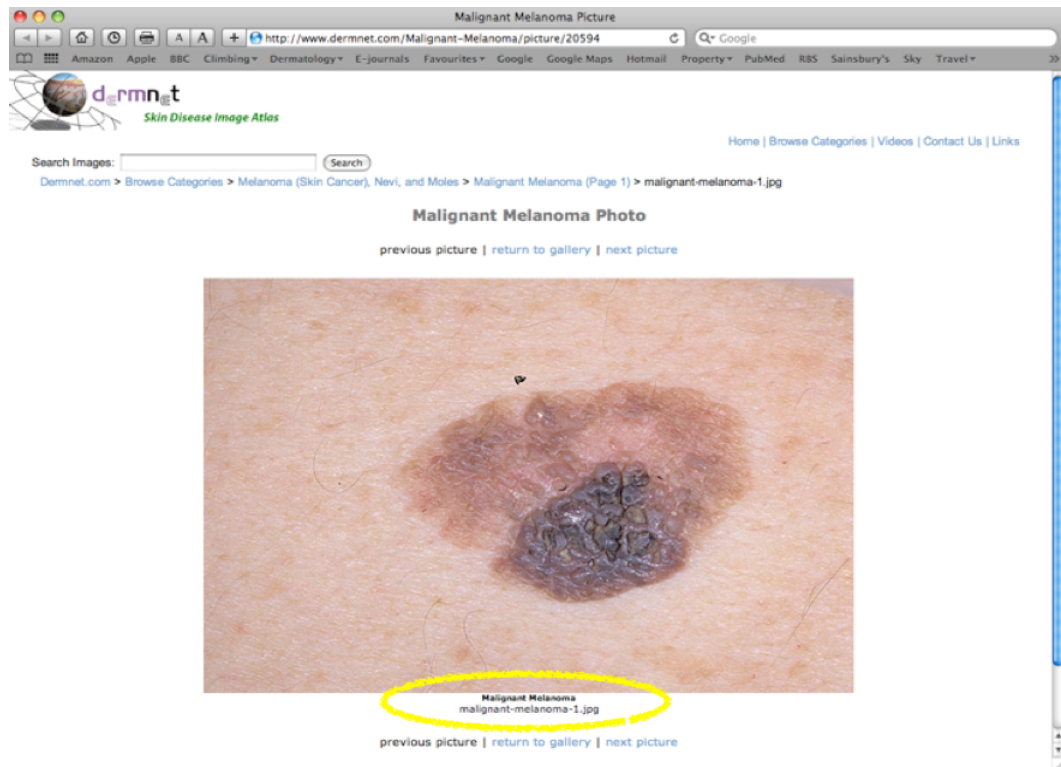


Figure 6.1b: Screenshot from Dermnet skin disease image atlas of the number one “melanoma” - this is the Internet’s largest public skin image bank (196).

In skin cancer education there is no experimental work directly comparing the effectiveness of learning on real patients versus learning from images. Provided images are effective as teaching tool, large databases of such images could address the main obstacle to effective dermatological teaching by removing the reliance on a constant supply of examples of new lesions. Furthermore, if the images were available online, widespread access could be achieved at relatively low costs, allowing standardized exposure across multiple institutions, thus limiting the intrinsic fluctuations that currently occur within different teaching organisations.

6.2 Aim

The two studies presented in this chapter investigate whether there is a role for clinical simulation, in terms of exposure to digital images of skin lesions without additional formal tuition, in the improvement of students’ diagnostic accuracy.

6.3 Materials & Methods

Subjects

Study 1:

All students (n=91) attending their two-week undergraduate dermatology attachment between May and August 2011 were asked to participate. In total six separate cohorts of between 14 and 16 students attended the Department over this time period. 57% (52/91) were female.

Study 2:

In study two all students (n=75) attending their two-week undergraduate dermatology attachment between July and September 2010 were asked to participate. These students attended the Department as five separate cohorts of between 13 and 17 individuals. 51% (38/75) were female.

Study 1: Immediate effects of simulated exposure

On the first morning of their attachment, before any formal instruction, all subjects undertook a 21-minute digital slideshow, in which they were asked to identify lesions from four multiple-choice options. This slideshow was conducted in addition to the regular teaching program with no other changes made to the students' dermatological attachment.

A major focus for non-experts when learning to identify cutaneous lesions is correctly diagnosing the three most common skin cancers and being able to readily distinguish them from their mimics: this study sought to assess the effect of exposure to teaching slideshows which were designed to address this real-life task.

Two 64-image separate test slideshows were constructed in a randomly stratified manner from 1323 suitable images in the Department's database. The first slideshow contained 64 "keratinocyte" (K) images, made up of sixteen images from each of the following four diagnoses BCC, SCC, actinic keratosis (AK) and intra-epithelial carcinoma (IEC); the second slideshow contained 64 "pigmented" images (P) made up

of sixteen images from each of the following four diagnoses Melanoma, dysplastic naevi (DN), benign naevi (BN) and seborrheic keratoses (SK). Within these eight categories all three of the common skin cancers (BCC, SCC, Melanoma) are represented, with what could be termed the main “pre-malignant” lesions (AK, IEC, DN) and their most commonly mistaken differentials (BN,SK). Furthermore the grouping of the lesions into “keratinocyte” and “pigmented” image batches was purposefully designed to be challenging, by placing the most commonly confused lesions together.

The order of the images was randomly allocated for both slideshows. Students were randomly allocated to either the keratinocyte slideshow (student cohorts A/B/C/D) or the pigmented lesion slideshow (student cohorts E/F). During these 64 image slideshows the students were required to identify each of the images by ticking one of four options on a purpose built multiple-choice answer sheet that only included the relevant four diagnoses. Each image was displayed without a diagnostic label for ten seconds, during which time they entered their answer on the multiple-choice sheet, immediately followed by the same image displayed with the correct diagnostic label for ten seconds so that they received instant feedback on their performance as they proceeded through the slideshow. Thus, the total duration of each of the slideshows was just over 21 minutes.

The overall accuracy of the student cohorts as they progressed through the slideshows was recorded. In addition, to ascertain whether individual student performance altered over the course of a slideshow, 24 of the 64 images were randomly selected and grouped into two batches, both consisting of twelve images (three images for each of the slideshows’ four diagnoses). These were placed as sequential images 5-16 (the initial test batch) and images 53-64 (the final test batch). Individual student’s accuracy was compared between the initial and final test batches to see if there had been an improvement in their performance over the course of the slideshow. The initial and final test batch images were reversed half way through the student cohorts to negate any difference in baseline difficulty between the two twelve image test batches.

Study 2: Dosing effects of simulated exposure

The students were shown two 12-minute slideshows, which were conducted in addition to the regular teaching program. No other changes were made to the students' dermatological attachment. The first of the additional slide shows was shown on day 1 (D1) of their two-week attachment and the second on day 6 (D6). Assessment of the effects of the slideshows was undertaken through a separate test slideshow on the final day (D10) of their attachment. Individual student attendance was recorded for each of the slideshows.

The D1 and D6 teaching slideshows both contained 72 diagnostically labelled images that were displayed for ten seconds each, giving a total teaching slideshow duration of twelve minutes for each. Three separate 72-image teaching slideshows and a single 36-image test slideshow were constructed in a randomly stratified manner. "Keratinocyte" and "pigmented" teaching slideshows were created as per study one.

To assess whether the teaching slideshows were facilitating the development of generic dermatological visual skills or more focused content specific skills, as a form of control four categories of "other" images were chosen to be visually distinct from the main lesions used in the teaching slideshows. This "other" or "control" teaching slideshow was made up of sixteen images from each of the following four diagnoses haemangiomas (H), pyogenic granulomas (PG), dermatofibromas (DF) and acrochordon/skin tag (A).

As the purpose of the study was to assess whether simulated exposure can improve student lesion identification (rather than factual student knowledge) a single example of each of the test diagnoses was provided in all teaching slideshows to ensure all students had been introduced to the appropriate terminology. The resultant four slideshows (control images, keratinocyte images, pigmented images and test images) were composed as per the table below – see Table 6.3a.

Slideshows	Diagnostic categories of lesions
Control/other images	16 x dermatofibroma (DF) 16 x haemangioma (H) 16 x pyogenic granuloma (PG) 16 x acrochordon/skin tag (A) & 1 of each of the 8 diagnoses from K & P slideshows Total 72 images
Keratinocyte images	16 x basal cell carcinoma (BCC) 16 x squamous cell carcinoma (SCC) 16 x actinic keratosis (AK) 16 x intra-epithelial carcinoma (IEC) & 1 of each of the 8 diagnoses from C & P slideshows Total 72 images
Pigmented images	16 x melanoma (Mel) 16 x dysplastic naevus (DN) 16 x benign naevus (BN) 16 x seborrheic keratosis (SK) & 1 of each of the 8 diagnoses from C & K slideshows Total 72 images
Test images	4 of each of the 4 K diagnoses (4 x BCC, 4 x SCC, 4 x AK, 4 x IEC) 4 of each of the 4 P diagnoses (4 x Mel, 4 x DN, 4 x BN, 4 x SK) & 1 of each of the 4 C diagnoses (DF, H, PG, A) Total 36 images

Table 6.3a: Table detailing the of the structure of the various teaching and test slideshows for Study 2.

The five cohorts of students were randomly assigned to different combinations of slideshows so that students underwent either single or double exposure to the control or teaching slideshows – see Table 6.3b.

Student Cohort	D1 teaching slideshow	D6 teaching slideshow
G (n=15)	Control/other images	Control/other images
H (n=15)	Pigmented images	Pigmented images
I (n=13)	Pigmented images	Control/other images
J (n=15)	Keratinocyte images	Keratinocyte images
K (n=17)	Keratinocyte images	Control/other images

Table 6.3b: Table showing the combination of teaching slideshows allocated to each student cohort in Study 2.

All slideshows were performed using Keynote (APPLE, California) on a 15" Macbook Pro (APPLE, California) connected via a HDMI link to a LG 55" HD TV (LG, South Korea).

On Day 10 the same 36-image slideshow was used to assess all the students to ensure standardization of test difficulty. As highlighted above, the test slideshow was composed of equal numbers of each of the pigmented and keratinocyte diagnoses. Such a construction allowed the effects of the two teaching slideshows to be assessed using a single test, as the overall test results could be sub-divided to examine:

- i) the accuracy for pigmented test images (n=16)
- ii) the accuracy for keratinocyte test images (n=16)

Four images of the other/control images were included to allow some measure of student performance for these. The students wrote their answers on purpose designed answer sheets and had 20 seconds to answer each of the test images. Again answers were "generously" accepted as correct even if they contained spelling mistakes, incomplete terminology, abbreviations and lay terms.

6.4 Results

Study 1: Immediate effects of simulated exposure

79 students completed their assigned slideshow, giving an enrolment rate of 87% (79/91) of students attached to the Dermatology Department during the study period. 54% of these students (43/79) were female. The combined overall accuracy of the students for the two slideshows was 49% (2499/5056) with a median score of 31 images of the 64 images correctly identified. There was no difference in accuracy between the sexes. There was also no difference in accuracy between the different student cohorts undertaking each of the test slideshows. The students' accuracy, however, did vary between the pigmented and non-pigmented slideshows at 45% (799/1792) and 52% (1700/3264) respectively. A Wilcoxon test showed this to be of significance ($p < 0.001$).

Combining the results for both the keratinocyte and pigmented teaching slideshows demonstrated that the students' overall diagnostic accuracy improved as they progressed through the slideshows, with a significant correlation ($Rho = 0.46$, $p < 0.001$) between the overall students' accuracy and the number of lesions to which they had been exposed.

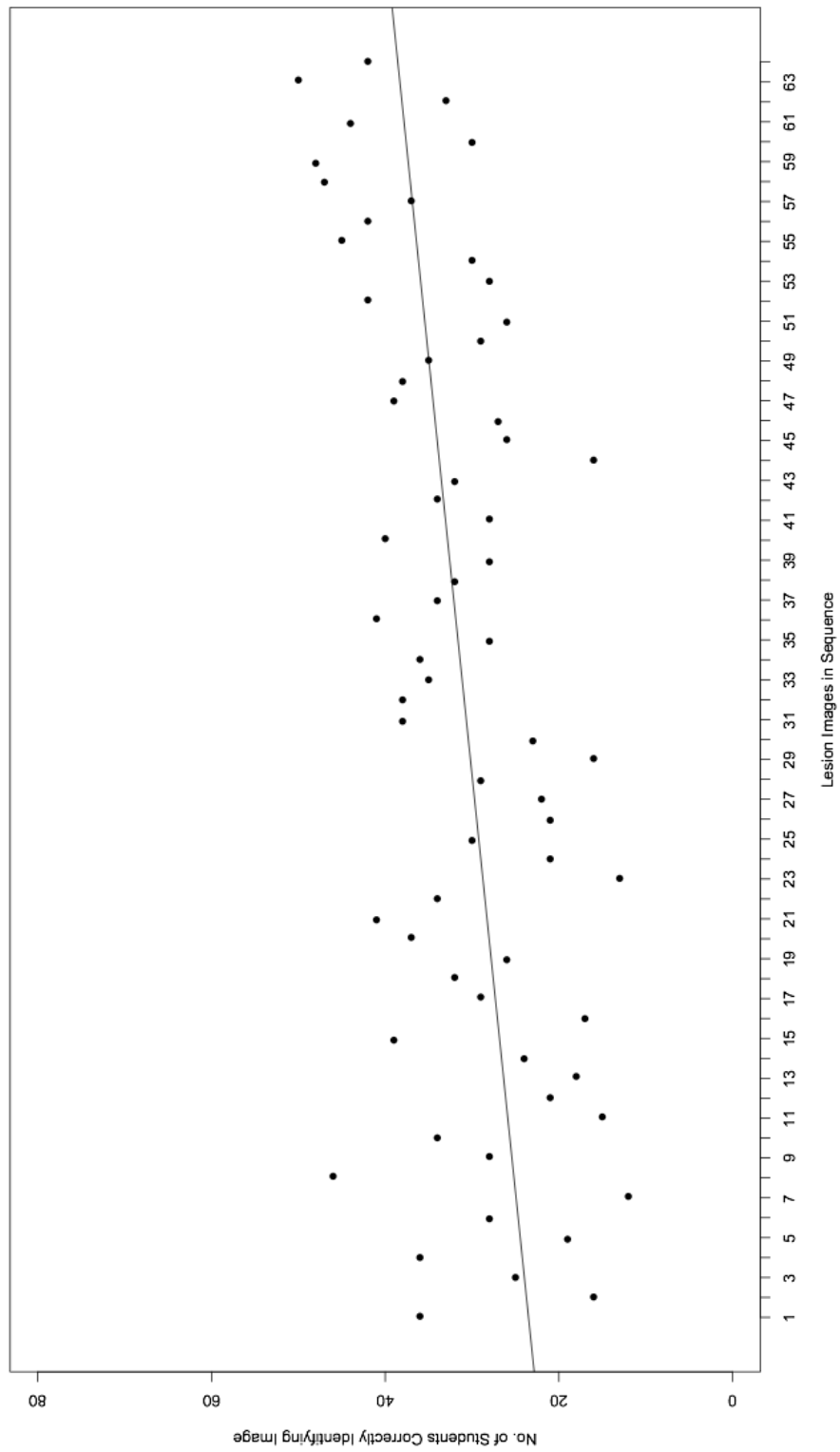


Figure 6.4a: Plot showing the students' overall combined diagnostic accuracy (pigmented and keratinocyte skin lesions) as they progressed through the 64-image slideshow.

A linear model of $y = 0.23x + 23$ ($p < 0.001$) was derived for the slideshows, indicating that during these slideshows for every 10 lesions, students' accuracy increased by 2.9%.

Individual student's performance was also shown to improve over the course of the slideshow with an accuracy of 44% (414/948) and a median score of 5 in the initial 12 test images, compared to an accuracy of 60% (567/948) and a median score of 7 for the final 12 test images. A Wilcoxon paired test confirmed that the differences in the students' scores between the start and end of the slideshow were significant ($p < 0.001$).

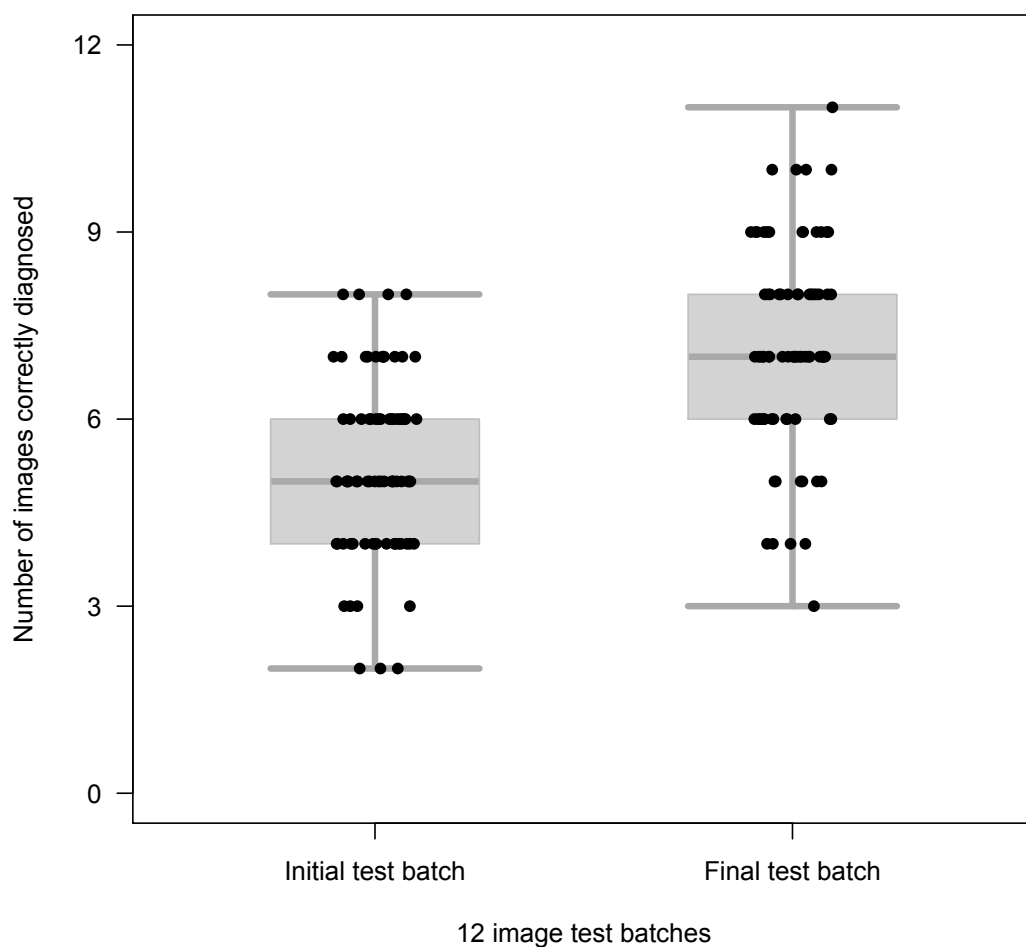


Figure 6.4b: Combined scatter and box plot showing the individual student's combined diagnostic accuracy (pigmented and keratinocyte skin lesions) for the initial and final twelve image test batches.

Keratinocyte test results

Cohorts A-D (n=51) who were exposed to the keratinocyte slideshow (K) had an accuracy of 44% (270/612) in the initial twelve test-images with a median score of 5, whereas for the final twelve test-images their accuracy was 65% (395/612) with a median score of 8.

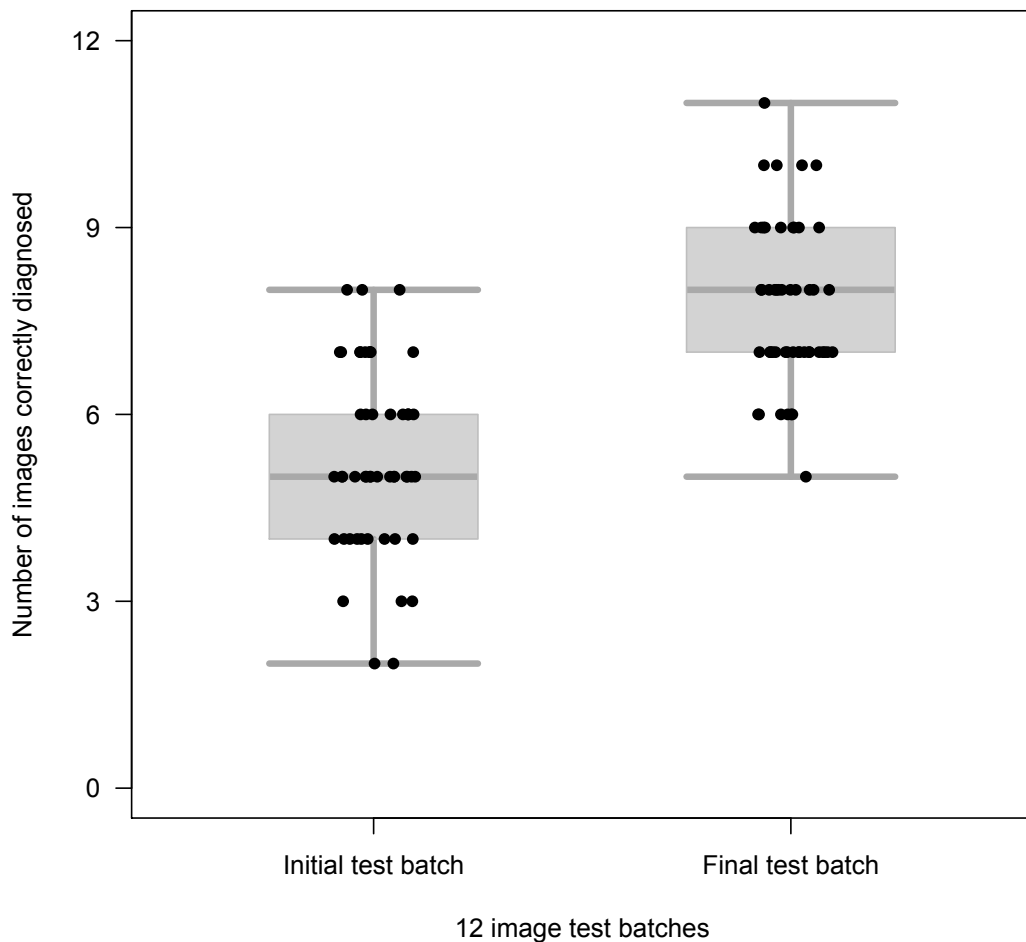


Figure 6.4c: Combined scatter and box plot showing the individual student's keratinocyte diagnostic accuracy for the initial and final twelve image test batches.

A Wilcoxon paired test confirmed that the differences in the students' scores between the start and end of the slideshows were significant ($p < 0.001$). Furthermore, there was a significant correlation ($Rho = 0.33$, $p < 0.01$) between the overall students' accuracy and the number of lesions to which they had been exposed.

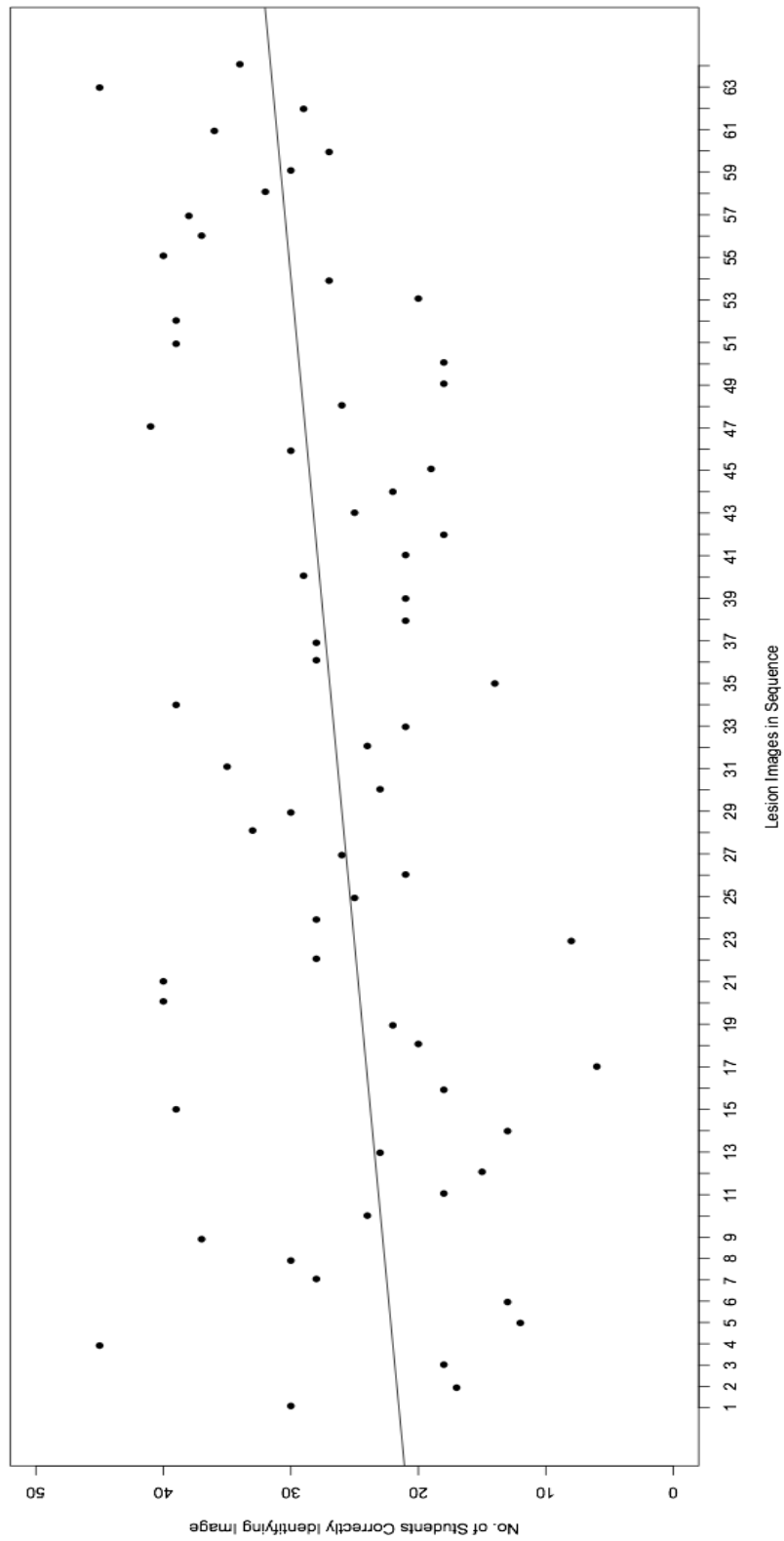


Figure 6.4d: Plot showing the students' overall keratinocyte diagnostic accuracy as they progressed through the 64-image slideshow.

Pigmented test results

Cohorts E&F (n=28) who were exposed to the pigmented slideshow (P) had an accuracy of 43% (144/336) in the initial twelve test-images with a median score of 5, whereas for the final twelve test-images their accuracy was 51% (172/336) with a median score of 6. Again, a Wilcoxon paired test confirmed that the differences in the students' scores between the start and end of the slideshows were significant ($p < 0.05$).

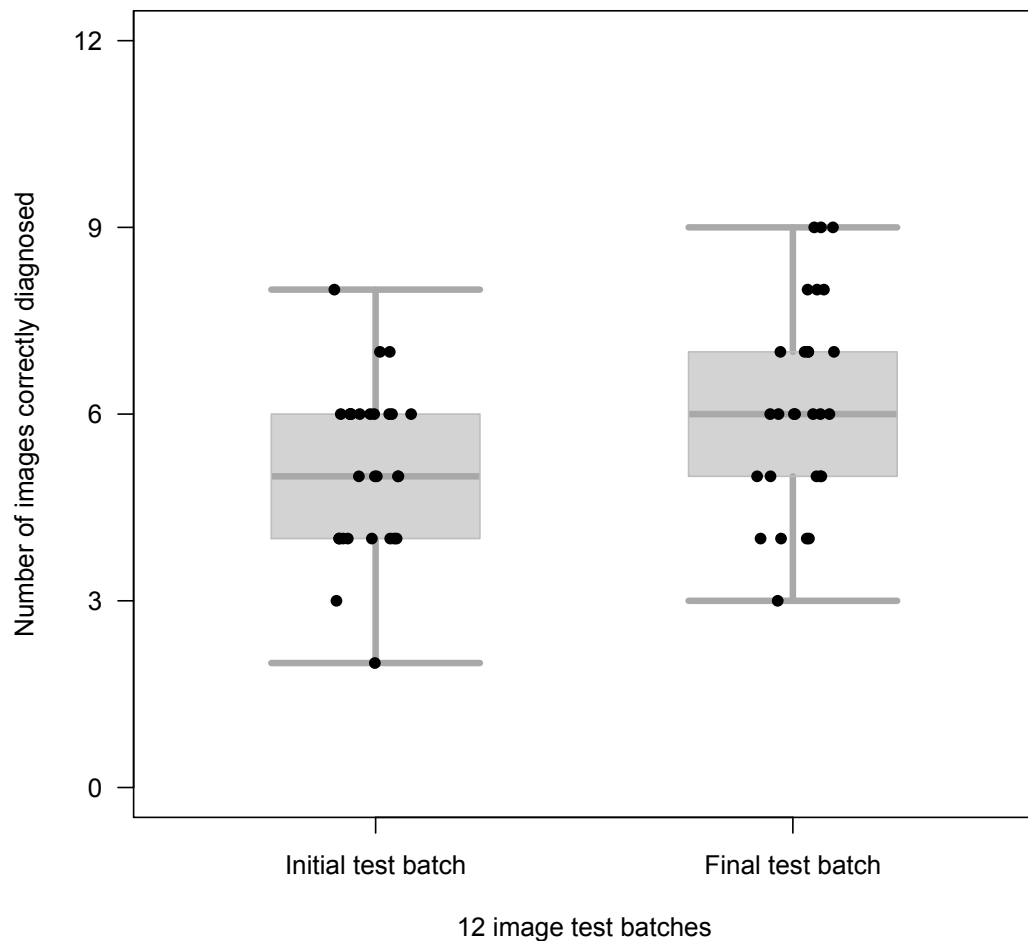


Figure 6.4e: Combined scatter and box plot showing the individual student's pigmented lesion diagnostic accuracy for the initial and final twelve image test batches.

In contrast to the keratinocyte slideshow, whilst there was a trend to a positive correlation ($Rho=0.15$) between the overall students' accuracy and the number of lesions to which they had been exposed, this did not reach significance.

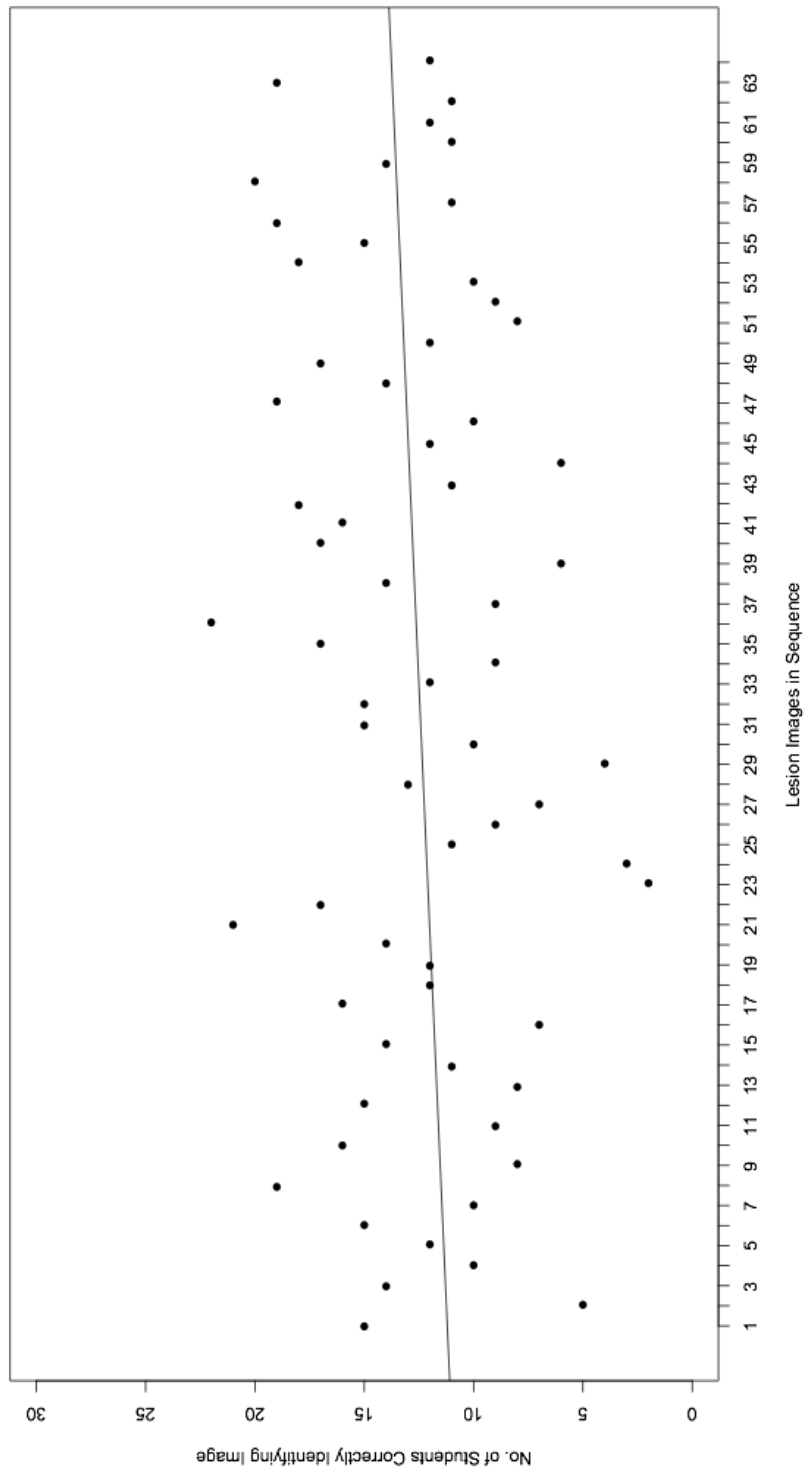


Figure 6.4f: Plot showing the students' overall pigmented lesion diagnostic accuracy as they progressed through the 64-image slideshow.

Study 2: Dosing effects of simulated exposure

Fifty-five students completed all three components of the study (their two allocated slideshows -on D1 and D6- plus the D10 test). All students that missed any of the components were excluded from further analysis (n=20). On this basis the study included 73% (55/75) of the students asked to participate. 56% (31/55) of these students were female. The students' overall accuracy in the D10 test was 38% (745/1980). There was no difference in overall accuracy between the sexes. The mean scores for all components of the test slideshow (overall score n=36, pigmented scores n=16, keratinocyte score n=16 and control score n=4) are displayed below, arranged by the different teaching strategies that to which each cohort of students was exposed.

Cohort	D1 teaching slideshow	D6 teaching slideshow	D10 test overall score (n=36)	D10 test pigmented score (n=16)	D10 test keratinocyte score (n=16)	D10 test control score (n=4)
G (n=7 of 15)	Control	Control	11	4.4	4	2.6
H (n=10 of 15)	Pigmented	Pigmented	14.6	8.1	6.1	0.4
I (n=11 of 13)	Pigmented	Control	12.4	6.8	3.6	1.9
J (n=13 of 15)	Keratinocyte	Keratinocyte	14.9	5.1	9.3	0.5
K (n=14 of 17)	Keratinocyte	Control	13.7	4.9	6.4	2.5

Table 6.4g: Table showing the students' mean scores on the test slideshow for each component, arranged according to the teaching slideshows undertaken.

There was also no significant difference in the overall test scores between the different cohorts of students exposed to the teaching slideshows (student cohorts H-K).

Combining the results for both the pigmented and keratinocyte teaching slideshows demonstrated that the students who were exposed to the 'teaching' slideshows twice had an accuracy of 55% (202/368) with a median score of 9 for the 'taught' diagnoses, compared to 41% (164/400) and a median score of 7 for the single exposure groups, the students who were not exposed to the tested images had an accuracy of 30% (294/992) with a median score of 5. A Kruskal-Wallis test confirmed these differences to be significant ($p < 0.001$). See Figure 6.4h below.

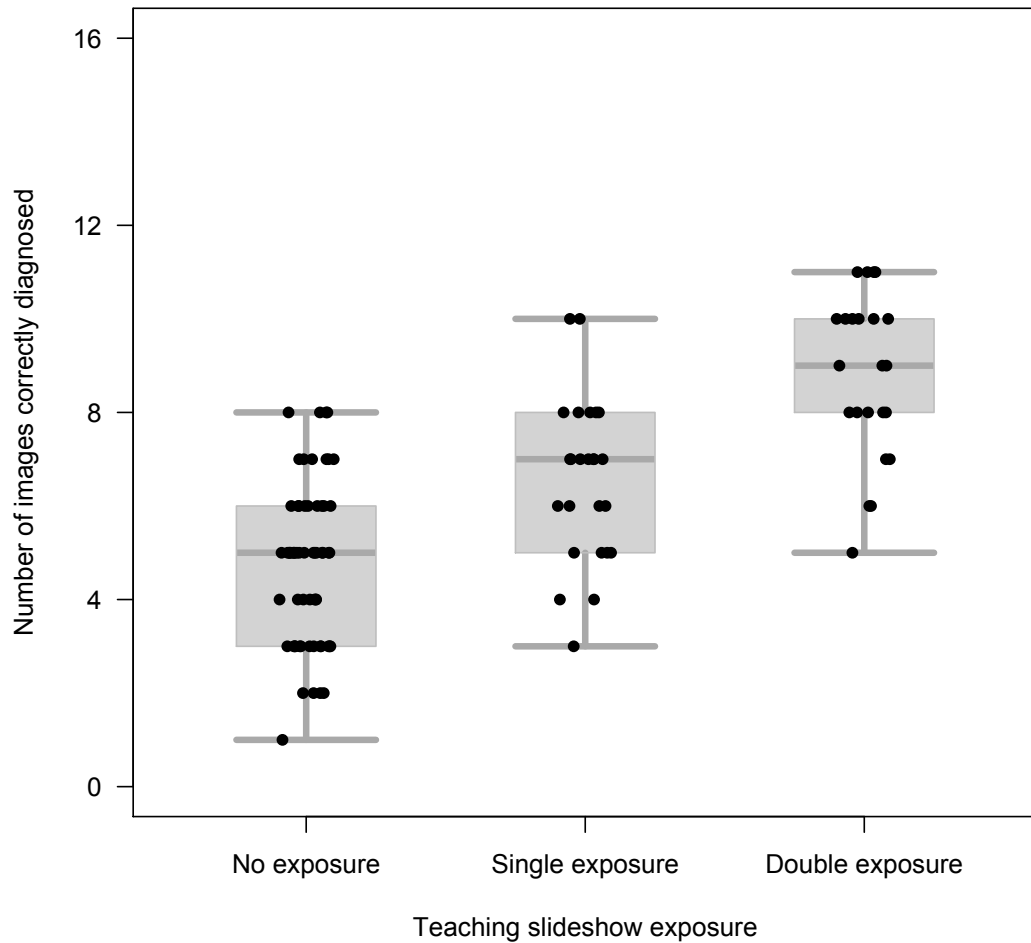


Figure 6.4h: Combined scatter and box plot showing the students' scores on the test slideshow, arranged according to their exposure to the teaching images.

Pigmented test results

Student cohort H (n=10), who were exposed twice to the pigmented teaching slideshows had a diagnostic accuracy for the sixteen pigmented test images of 51% (81/160), with a median score of 8. Student cohort I (n=11), who were exposed to a single pigmented teaching slideshow had a diagnostic accuracy of 43% (75/176), with a median score of 7. For student cohorts G, J & K (n=34) that had not been exposed to a pigmented teaching slideshow their diagnostic accuracy was 30% (165/544), with a median score of 5. See Figure 6.4i below.

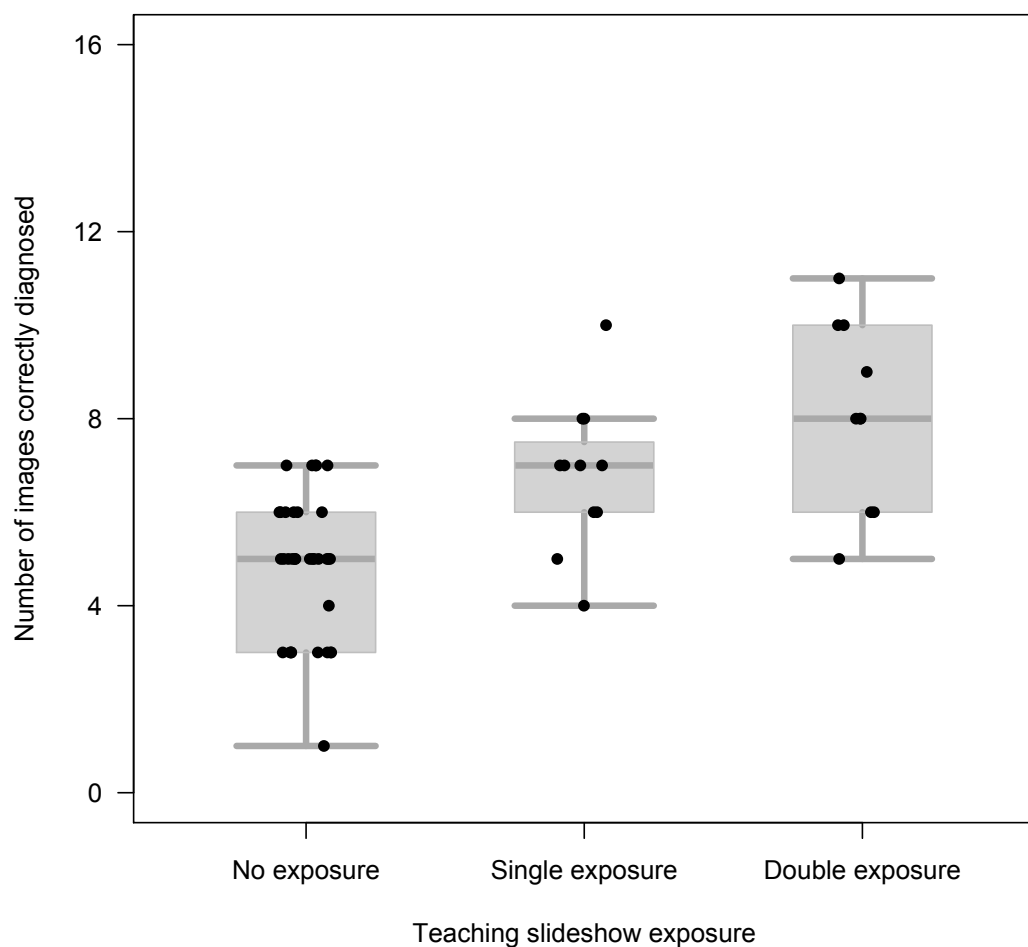


Figure 6.4i: Combined scatter and box plot showing the students' scores for the pigmented image component of the test slideshow, arranged according to their exposure to the teaching images.

A Kruskal-Wallis test confirmed these differences to be significant ($p < 0.001$). Furthermore, the improvement in the students' accuracy was seen across all four of the pigmented lesion diagnoses tested. See Table 6.4j below.

Students' accuracy for pigment lesion diagnoses, grouped by exposure to pigmented teaching slideshows			
Diagnosis	No exposure (Cohorts G,J,K) n=34	Single exposure (Cohort I) n=11	Double exposure (Cohort H) n=10
Melanoma (Mel)	50% (68/136)	57% (25/44)	58% (23/40)
Dysplastic naevus (DN)	4% (6/136)	34% (15/44)	45% (18/40)
Benign naevus (BN)	53% (72/136)	55% (24/44)	63% (25/40)
Seborrheic keratosis (SK)	14% (19/136)	25% (11/44)	38% (15/40)
Combined	30% (165/544)	43% (75/176)	51% (81/160)

Table 6.4j: Table showing the students' diagnostic accuracy for each of the four diagnoses contained within the pigment lesion component of the test.

Keratinocyte test results

Student cohort J (n=13), who were exposed twice to the keratinocyte teaching slideshow had a diagnostic accuracy of 58% (121/208), for the sixteen non-melanoma skin cancer test images, with a median score of 10. Student cohort K (n=14), who were exposed to a single keratinocyte teaching slideshow had a diagnostic accuracy of 40% (89/224), with a median score of 6.5. For student cohorts G, H & I (n=28) who had not been exposed to a keratinocyte teaching slideshow their diagnostic accuracy was 29% (129/448), with a median score of 4. See Figure 6.4k below.

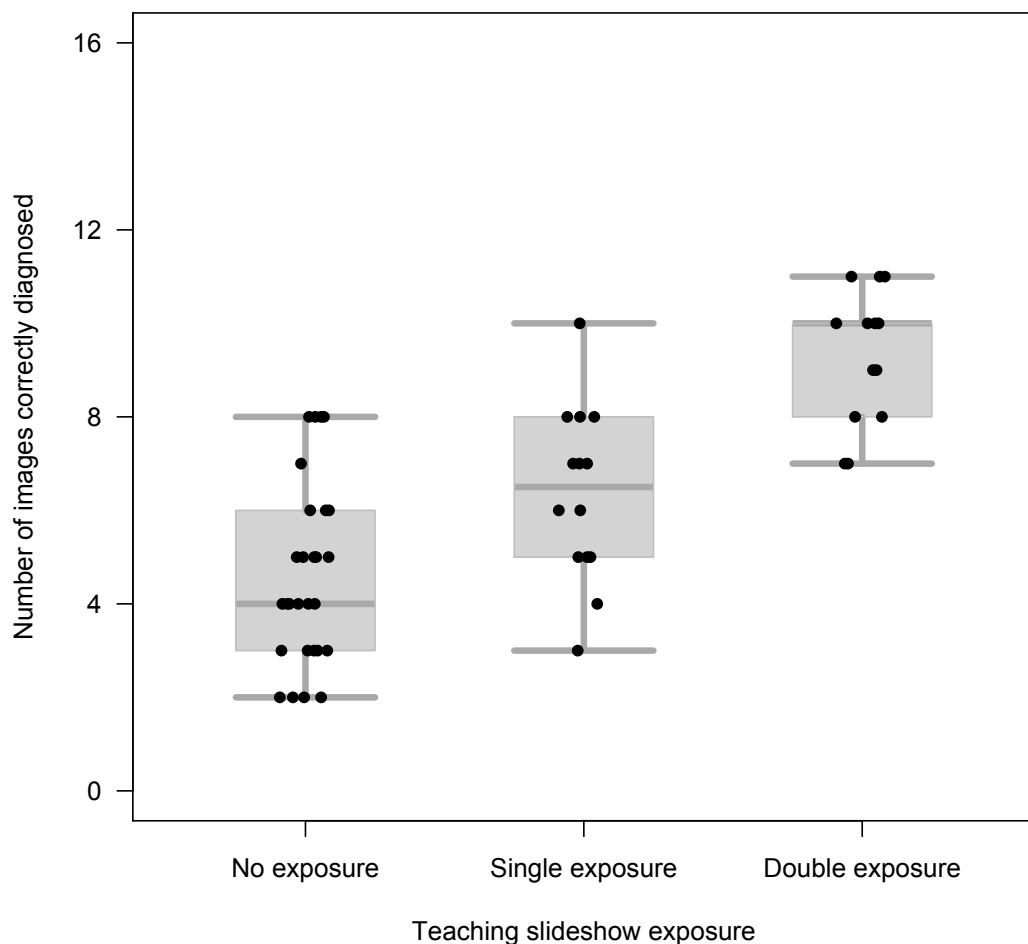


Figure 6.4k: Combined scatter and box plot showing the students' scores for the keratinocyte image component of the test slideshow, arranged according to their exposure to the teaching images

A Kruskal-Wallis test confirmed these differences to be significant ($p < 0.001$). Again, the improvement in the students' accuracy was seen across all four non-melanoma skin cancer diagnoses tested. See Table 6.4l below.

Students' accuracy for non-melanoma skin cancer diagnoses, grouped by exposure to keratinocyte teaching slideshows			
Diagnosis	No exposure (Cohorts G,H,I) n=28	Single exposure (Cohort K) n=14	Double exposure (Cohort J) n=13
Basal cell carcinoma (BCC)	30% (34/112)	36% (20/56)	56% (29/52)
Squamous cell carcinoma (SCC)	41% (46/112)	57% (32/56)	69% (36/52)
Actinic keratosis (AK)	38% (42/112)	50% (28/56)	71% (37/52)
Intra-epithelial carcinoma (IEC)	6% (7/112)	9% (5/56)	37% (19/52)
Combined	29% (129/448)	40% (89/224)	58% (121/208)

Table 6.4l: Table showing the students' diagnostic accuracy for each of the four diagnoses contained within the keratinocyte lesion component of the test.

A graphical display of the combined data of both tables above, demonstrates that whilst there is a baseline difference in accuracy of students for different diagnostic classes of lesions, across all the classes there is improvement in accuracy with increased exposure. See Figure 6.4m below.

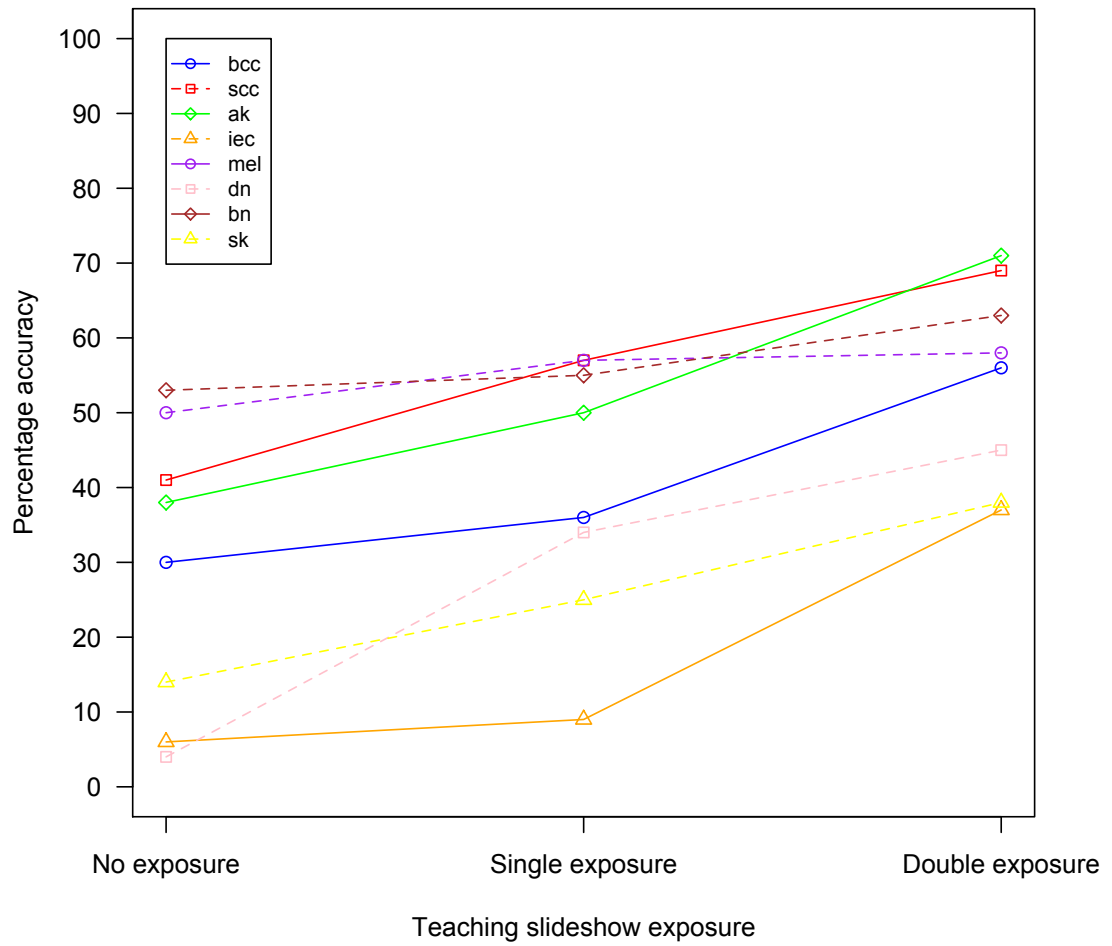


Figure 6.4m: Plot showing the students' learning curves for the identification of cutaneous malignancies, arranged by individual diagnostic class and according to their exposure to the teaching images. Abbreviations are as per Tables 6.3a, 6.4j, 6.4l.

6.5 Discussion

Both studies unequivocally demonstrate that when tested on similar images, students exposed to clinical simulation, in the form of labelled digital images have developed significantly improved diagnostic accuracy. As this improvement resulted from relatively limited exposure (10-20 seconds per image) and without any additional formal tuition the most likely explanation is that the simulation facilitates the students acquisition of NAPR.

The students' absolute accuracy resulting from the clinical simulation is not critical because, as highlighted previously, such measures only reflect the relative difficulty of any given test. It was therefore not surprising that the accuracy of the students was notably higher in the first study (49% vs 38%) given this only required an answer from four multiple-choice options, meaning by chance alone a baseline accuracy of 25% might be expected.

The studies did show that the students' identification of lesions was both content specific and exposure dosage responsive. Exposure dosage being a composite of number of images and reinforcement by repeated simulation. The use of 'control' images in the second study did not improve the students' diagnosis of the images tested. The slideshows only seemed to have an effect when relevant images were displayed. This finding is reinforced by the improved performance in the students undergoing double exposure to the slideshows. Whilst this is perhaps unsurprising given the premise that students are unconsciously developing NAPR, it contrasts with some of the published medical education literature where the effects of content specificity are felt to be limited (197).

Furthermore, both studies demonstrated that the rate at which students' accuracy improves varies depending on the specific lesion class: the learning curves for diagnosing cutaneous lesions are not uniform. In particular, it would seem the students found differentiating pigmented lesions to be more difficult. In the first study the students' overall accuracy on the pigmented diagnoses was significantly lower (45% vs 52%) and the correlation with improved accuracy as they progressed through the

slideshow weaker (Rho 0.33 vs 0.15). In the second study the learning curves for the pigmented diagnoses are flatter and there was a trend towards higher accuracy in the diagnosis of keratinocytic lesions. Some of this variance in the learning curves could be due to the design of the studies; in that by grouping together commonly confused lesions the task was made purposively challenging. Although challenging, this design perhaps better reflects the difficulties faced in clinical practice; a lay person themselves may be able to distinguish between a 'scaly' or 'brown' lesion, but within these simple classifications there remain multiple differentials for a clinician to delineate in the patient presenting with a new lesion. The most obvious explanation for the different rates of diagnostic improvement is that the visual overlap between the four pigmented diagnoses was greater than that for the four keratinocyte diagnoses, thus making discrimination more difficult.

Across the eight diagnoses tested the lowest level of diagnostic improvement was seen for melanomas and benign moles (<10% improvement in both during study two), further reflecting the students' relative difficulty with pigmented lesions. This contrasts with the observation that for these two diagnoses the students had the highest baseline accuracy. The students baseline accuracy (i.e. accuracy with no teaching slideshow simulation) must reflect their peri-attachment knowledge and experience, and although this will vary from student to student (for which there is no adequate control) the results demonstrate that the commonly emphasized diagnoses had higher baseline accuracies.

The studies provide some insight into just how many lesions are required to substantially improve diagnostic abilities. Whilst sixteen images of specific lesional diagnoses may have been sufficient to witness a statistical improvement in students' score, the resulting overall accuracy is nowhere close to being acceptable for independent practice. Although sixteen training examples may be considered modest it is worth reiterating that no student witnessed sixteen 'in vivo' examples of any individual diagnoses in the studies described in the previous chapter. The pigmented teaching slideshows exposed the students to sixteen melanomas- far more than is achievable during a standard attachment. Indeed, back of the envelope calculations would suggest that even for a dermatology trainee sixteen melanomas are likely to

represent half what they see in their specialty training in South East Scotland (based on the assumptions - 80 melanomas/year, four years training, standard leave protocol and experience diluted by an eight person rota).

If the linear model derived from the combined results of study one is extrapolated it suggests that to achieve 95% accuracy would require the students to have been exposed to 57 examples of each lesion. Obviously, such a generalization is flawed as the model is derived from multiple choice answers and extending the model presumes that there is no plateau to the learning curves and disregards the fact that not all lesions are equally difficult. This extrapolation simple provides a rudimentary indication of the number of images of lesions that may be required to develop diagnostic competence. For the eight common lesions used within these studies around 500 training examples might be required. Given these eight diagnoses reflect only a small spectrum of the total number of possible cutaneous lesional diagnoses the true number of example lesions required to develop diagnostic competence is likely to run into the thousands.

Whilst there is little doubt that practical experience increases diagnostic aptitude, developing teaching tools to improve diagnostic competence for cutaneous lesions is not straightforward. Although these studies demonstrate that simulation is likely to have a role in improving novices' (and indeed trainee dermatologists') identification of skin cancers further studies are indicated to identify where, if anywhere, the diagnostic benefit of spending more time on exposure to such images plateaus.

Chapter 7

3D or not 3D?

The role of depth perception on skin lesion diagnosis

7.1 Introduction

Some of the work presented in this chapter has been published in Arch Dermatol. 2010; 146(10):1184-5; PMID: 20956667. "Teaching dermatology using 3-dimensional virtual reality."

Many dermatologists believe that the three-dimensional (3D) properties of cutaneous disease (rather than the 2D afforded by photographic representations) have an important role in achieving the correct diagnosis. For this reason up until the early 20th Century wax moulages were considered the gold standard educational tool as their combination of both realistic colourings and 3D depth afforded the best available surrogates to real-life examples for teaching (198,199).



Figure 7.1a: Photograph of traditional wax moulage showing seborrheic keratoses (verrucae seniles) (198).

With the advent of affordable colour photography, the inherent costs of producing and transporting such moulages meant their use as educational resources diminished greatly (200). Recently, however, there has been resurgence in the use of modern prosthetic moulages for training and testing medical students (201-206). It has been suggested these 'modern moulages' could be superior for teaching to conventional 2D photographs but robust experimental evidence to substantiate this claim is limited.

The modern imaging techniques that have been developed as part of the Dermofit project allow for accurate 3D models of skin lesions to be captured simply and efficiently. These models have significant advantages over conventional 2D photography as they allow students to rotate and pan around the images as they would in real-life.



Figure 7.1b: Screen snapshot demonstrating one of the 3D models - a keratoacanthoma of the upper lip.

Although it may be reasonable to postulate that 3D image capture is likely to have additional merit for teaching and gaining vicarious clinical experience, it should not be assumed that merely because it involves novel technology that it is better than the traditional 2D photographic images. Indeed, one investigator has claimed that for

teaching anatomy it is difficult to demonstrate an advantage of 3D over 2D image models (207).

7.2 Aim

To examine the hypothesis that if 3D image models contain more diagnostically useful information than 2D photographs (of the same lesions), diagnostic accuracy should be higher when assessing the former than the latter.

7.3 Materials & Methods

Subjects

All medical students that completed their two-week undergraduate dermatology attachment between July 2010 and January 2011 were invited to participate. In total 119 students were enrolled (representing 84% of eligible students). Fifty-nine were female (50%). None of the students had prior exposure to any of the 3D models during their respective attachments.

Image selection

One hundred images were randomly selected in a stratified manner from 687 suitable images in the Dermofit image library. These 100 images consisted of 20 images from five common lesion classes; actinic keratoses, basal cell carcinomas, benign naevi, squamous cell carcinomas and seborrhoeic keratoses. All images had been captured using the same controlled fixed distance photographic setup described fully in Chapter two (Generic Methods). Briefly a pair of conventional 2D digital colour images (JPEG format) were taken for each lesion from which 3D models were generated using Dimensional Imaging's custom built software (Dimensional Imaging, Glasgow). The 2D digital colour image from the camera directly perpendicular to the lesion was used as the corresponding 2D image for the study.

Primary study

Each subject was required to provide written diagnoses for a different selection of 40 test images. These 40 test images were randomly sampled (from the original 100 images) in a stratified manner so that each student was tested on two sequential batches, each of 20 different images (with both test batches consisting of four examples of each of the five lesion classes). For each subject the 40 images were displayed on the same Apple (APPLE, California) iMac computer using MATLAB software (MATHWORKS, Massachusetts) under similar ambient lighting conditions. The first 50 students' diagnostic accuracy was assessed on twenty 2D images followed by twenty 3D models. The students were not formally taught on any of these images or given feedback until the conclusion of the Study. The study was designed to ascertain whether there was a difference in their accuracy between the two image types.

Additional studies

To ensure that the results of the primary experiment reflected the principal hypothesis being tested, five additional experiments were devised both to confirm the experimental findings and importantly to exclude any confounding factors arising from the experimental design. The remaining 69 students undertook the following additional studies.

Additional study 1: To confirm the experimental findings thirteen students undertook the same experimental protocol as the original 50 students, with their 40-lesion test consisting of a first batch of twenty 2D images and second batch of twenty 3D models.

Additional study 2: To investigate whether an improvement across the 40-lesion test was observed as students became accustomed to the testing process, thirteen students were tested on two sequential batches that both consisted of twenty 2D images.

Additional study 3: To further assess whether the findings of the primary study were batch order dependent, fourteen students were examined by reversing the protocol of the primary study (these students were therefore first tested on a batch of twenty 3D models and then on a second batch of twenty 2D images).

Additional studies 4 & 5: As highlighted in the generic methods, one of the potential benefits of the 3D models is in addition to depth users can interact with the models as they might in real-life by ‘panning around’ and ‘zooming in’. These two additional studies sought to investigate whether the students achieved a better score simply because of these features. It is possible that such interactions could, by themselves, lead the students to focus more intently on the 3D models, and therefore any improvement in accuracy might actually be independent of the models additional 3D depth data. To investigate the effects of image interaction, in additional study 4, fourteen students were tested on a first batch of twenty 2D images followed by a second batch of another twenty 2D images for which it was possible to rotate the image to view it from multiple angles. Furthermore in additional study 5, fifteen students followed a similar protocol but for their second test batch instead of being able to rotate the twenty 2D images they were able zoom in on the images. For ease of reference these additional studies are summarised in Table 7.3a.

Additional Study	No. Students	1st test batch	2nd test batch	Reason for study
1	13	20 2D images	20 3D images	Assess repeatability
2	13	20 2D images	20 2D images	Assess effect of order
3	14	20 3D images	20 2D images	Assess reversibility
4	14	20 2D images	20 2D images (able to rotate 2 nd batch of images)	Assess effect of interaction
5	15	20 2D images	20 2D images (able to zoom in 2 nd batch of images)	Assess effect of interaction

Table 7.3a: Table outlining the five additional studies that the 69 students undertook.

Student feedback

Feedback was not provided until the students had completed their full 40-image tests. On completion of the feedback session the students that had been exposed to the 3D

models were asked to subjectively evaluate their experience using an anonymous three-point yes/no questionnaire: i) did you find the 3D images easy to 'navigate'? ii) did you find it easier to diagnose the 3D images? iii) do you think the 3D images might be better than 2D images for teaching?

7.4 Results

Primary study

The 50 students had an overall diagnostic accuracy of 38% (385/1000) with a mean score of 7.7 (median 8, range 2-12) for the 20 2D images and an accuracy of 47% (470/1000) with a mean score of 9.4 (median 9.5, range 5-14) for the 20 3D images. The differences between the students' accuracy were significant (Wilcoxon paired $p < 0.0001$). The results for these 50 students are displayed graphically below in Figure 7.4a. There was no difference in diagnostic accuracy between the sexes or between the individual student cohorts.

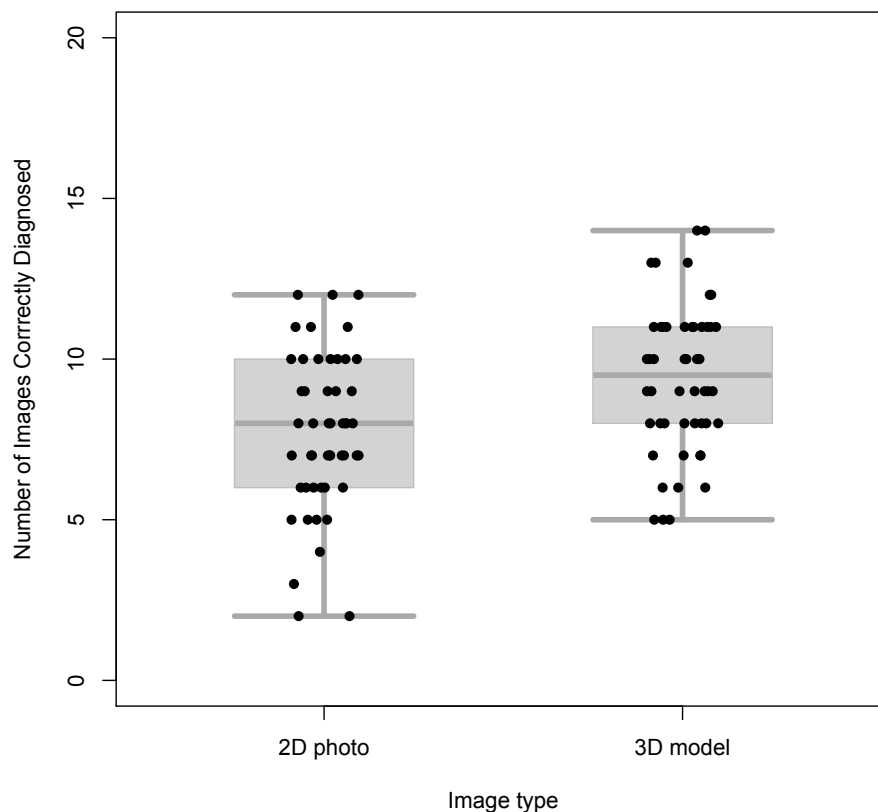


Figure 7.4a: Combined scatter and box plot showing the subjects' overall scores for diagnostic accuracy on the 2D photographs and the 3D models, in the primary study.

Additional studies

The results for the 69 students enrolled in the five additional experiments are displayed in the table below- see Table 7.4b.

Additional Study	Median Score 1 st test batch	Overall Accuracy 1 st test batch	Median Score 2 nd test batch	Overall Accuracy 2 nd test batch	Paired Wilcoxon
1	5 (3-10)	29% (77/260)	8 (4-12)	39% (103/260)	P<0.01
2	6.5 (1-10)	29% (82/280)	6 (2-9)	29% (84/280)	NS
3	8 (5-12)	41% (109/260)	6 (4-10)	32% (85/260)	P<0.01
4	5 (1-10)	27% (78/280)	6.5 (2-9)	29% (84/280)	NS
5	7 (2-10)	33% (99/300)	6 (4-10)	34% (104/300)	NS

Table 7.4b: Table outlining the results of the five additional studies.

Additional study 1 confirmed the results to be replicable. Additional studies 2 and 3 demonstrated that the improvement observed in accuracy was not due to batch order; as when the 3D model batches were shown after the 2D images, the same findings were observed as when the batch order was reversed (study 3) and no trend towards improvement was seen in the second batches (study 2). Additional studies 4 and 5 suggested that the improvement in accuracy was not due to a simple interaction effect as the actions of zooming in or rotating the images failed to improve the students' diagnostic accuracy.

Student feedback

For the 76 students that had been exposed to the 3D models; 96% (73/76) felt the 3D models were easy to navigate, 81% (62/76) felt the 3D models were easier to diagnose than the 2D images and 96% (73/76) felt the 3D models would be superior to 2D images for teaching.

7.5 Discussion

This study demonstrates that students can achieve the correct diagnosis more frequently for a range of common skin lesions when shown 3D images rather than 2D images. The absolute differences are modest, but are of course dependent on the difficulty of the identification task. The additional experiments reported suggest that this result is not an artefact due to secondary factors such as the level of interaction with the image (due to zooming or rotation of the image) varying between the experimental arms. It is worth noting the students preferred 3D images to 2D images but this of course may reflect their relative novelty.

It is important to emphasise that improved student learning was not shown with the use of 3D rather than 2D images, as this was not the immediate goal of the study. Rather, the results show experimentally that in this particular clinical skill domain, 3D images provide more diagnostically useful information than 2D images. Given the long history of moulage in dermatology it is tempting to conclude that these findings are unsurprising, however, as highlighted above the benefits of 3D images in teaching undergraduate anatomy have not been universally demonstrated (207). Although in other visual disciplines in medicine, such as radiology, more detailed images are widely acknowledged to be diagnostically superior.

The findings are also of importance for dermatological diagnosis by telemedicine. For financial and logistical reasons the use of teledermatology is becoming increasingly widespread and 3D imaging in telemedicine can (at a cost) be made available. In a thesis exploring ways in which identification of skin cancers can be enhanced, these preliminary findings suggest that further studies are merited investigating using 3D imaging in telemedicine diagnostic sessions.

Chapter 8

To what extent may “classical” clinical images inhibit the acquisition of diagnostic skills for skin lesions

8.1 Introduction

In a visual literate subject such as dermatology exposure to images are of inestimable value in acquiring diagnostic skill. The importance of visual images was appreciated in the earliest dermatology texts, even when reproduction of images could only be achieved at significant costs.

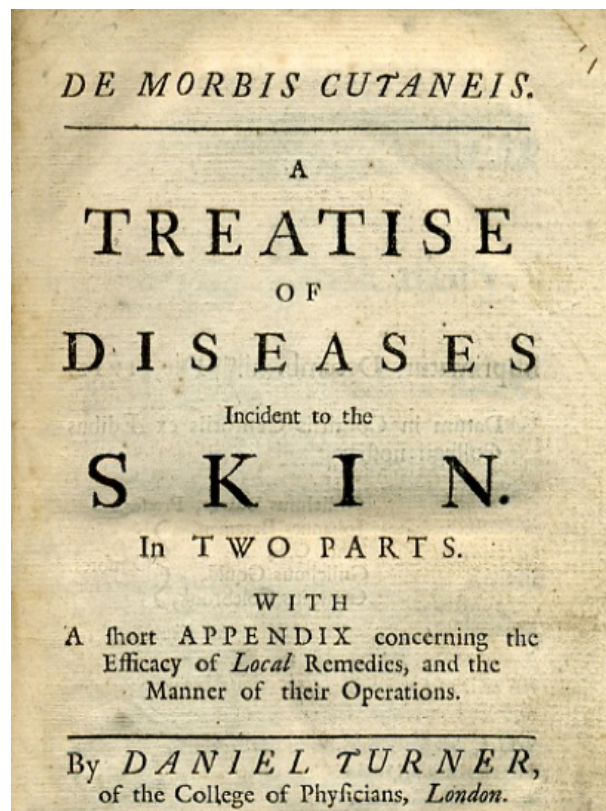


Figure 8.1a: Title page of Daniel Turner's "De morbid cutaneis. A treatise of diseases incident to the skin". Image courtesy of Professor J L Rees.

In the first English language textbook of dermatology “De morbid cutaneis. A treatise of diseases incident to the skin” by Daniel Turner, published in 1726, the sole illustration was that of the author, but such paucity of illustration is likely to have reflected more the costs of reproducing appropriate images, rather than any desire of the author for personal aggrandisement.

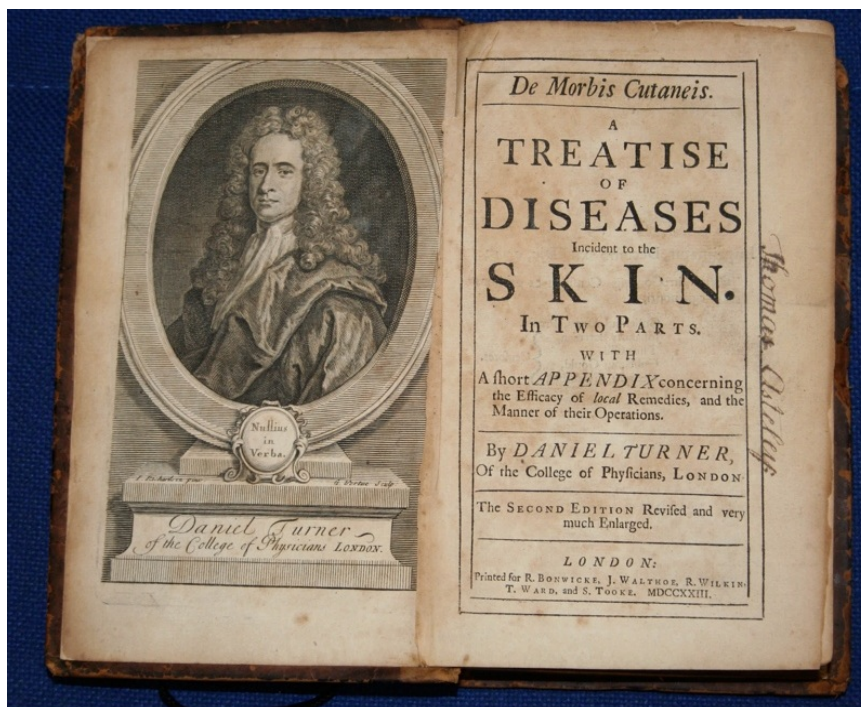


Figure 8.1b: Title page of Daniel Turner’s book alongside the textbook’s only illustration - the author himself. Image courtesy of Professor J L Rees.

The quandary confronting Turner is that which has confronted all authors of textbooks of dermatology namely assessing the cost/benefit of providing appropriate images. In consequence, the illustrations chosen tend to be those of the “classic” lesion or disease. As it is rare for a disease to manifest with all “classical” features, only with experience does the expert diagnostician come to appreciate the multiple possible variations with which a disease may manifest. Until the dawn of the electronic age it has therefore been impracticable for a publisher to provide the number of images that would be required to illustrate the many variants of any given skin disease. In this chapter, the possibility that exposure to standard textbook “classical” images of dermatological lesions may inhibit as well as improve diagnostic acumen is explored,

extending the work in previous chapters examining the benefits to be gained from exposure to electronic imaging through examples of non-classical images.

8.2 Aim

To assess and compare the diagnostic accuracy of students (at the end of their dermatology attachment) between skin lesion images reproduced from textbooks to those of lesions taken sequentially from a rapid access skin lesion clinic.

8.3 Materials & Methods

Subjects

Study 1:

All students (n=62) attending their two-week undergraduate dermatology attachment between April and June 2011 were asked to participate. These students attended the Department as four separate cohorts of between 15 and 16 individuals. 55% (34/62) of these students were female. On the final day of their attachment the students undertook a 60-image slideshow to test their diagnostic accuracy for pigmented lesions. No other changes were made to the students' dermatological attachment. The students were not given prior warning of the slideshow but were advised at the start of the test that the marks did not count towards their assessment.

Study 2:

A cohort of students (n=15) undergoing their dermatology attachment between October 2010 and February 2011 were asked to participate. In addition, all the dermatology specialty trainees in the Department of Dermatology over this period were also asked to participate (n=8).

Textbook image library

A library of "textbook" images was obtained by scanning all the images of the sixteen lesional diagnoses recommended in the BAD undergraduate curriculum from fifteen popular dermatological textbooks (157). The choice of the reference textbooks was

based on our previous survey of student resource utilization along in Chapter 5 and supplemented with the most popular current dermatological textbooks identified by online searches of both Amazon and Blackwell booksellers' websites using the key word "dermatology". A full list of the fifteen reference textbooks used is provided in Table 8.3a below.

Textbook	Lead Editor/Author	Edition
Davidsons principles and practice of medicine (168)	Boon	2006
Clinical dermatology (169)	Weller	2008
Dermatology: an illustrated colour text (170)	Gawkrodger	2007
Oxford handbook of clinical specialties (171)	Collier	2009
ABC of dermatology (172)	Buxton	2009
Crash course dermatology (173)	Cheung	2009
Physical signs in dermatology (174)	Lawrence	2001
Fitzpatrick dermatology in general medicine (175)	Wolff	2007
Differential diagnosis in dermatology (176)	Ashton	2004
Rook's textbook of dermatology (177)	Burns	2010
Lecture notes in dermatology (178)	Graham-Brown	2006
Fitzpatrick's colour atlas of clinical dermatology (208)	Wolff	2009
Dermatology (209)	Bolognia	2007
Atlas of clinical dermatology (210)	du Vivier	2002
Braun Falco's dermatology (211)	Burgdorf	2008

Table 8.3a: Table showing the fifteen textbooks used for generating the "classical" or "textbook" batch of images.

The relevant images were scanned at high resolution (1600 Megapixels) using an Epson Perfection 3200 PHOTO scanner (EPSON, Japan) and stored as non-compressed TIFF files. In total, 286 images were collected and categorised according

to their textbook diagnosis. The make-up of the 286 images by their diagnoses is listed below in Table 8.3b.

Textbook library		
Diagnoses	Number of images (n=286)	Percentage
Basal cell carcinoma	38	13%
Squamous cell carcinoma	22	8%
Melanomas/Melanoma in Situ	52	18%
Viral warts	12	4%
Cysts	6	2%
Melanocytic naevi	54	19%
Seborrhoeic keratoses	18	6%
Actinic keratoses	9	3%
Bowen's disease	10	3%
Dermatofibromas	11	4%
Keratoacanthomas	16	6%
Lipomas	5	2%
Pyogenic granulomas	9	3%
Mycosis fungoides	16	6%
Paget's disease	8	3%
Cutaneous metastases	0	0%

Table 8.3b: Table showing the diagnostic subcategories of the textbook library images.

Real-life image library

A library of “real-life” images was obtained by sequentially photographing the referral lesion(s) on all patients who attended a fortnightly skin lesion clinic at a district general hospital over a 9-month period, between January and October 2010. An 87% recruitment rate was achieved and the lesions were photographed as described in Chapter 2 and 10. Again, all images that matched one of the BAD's recommended

sixteen lesional diagnoses were selected (157). In total 283 images met the above criteria and the exact diagnoses of the lesions is listed below in Table 8.3c.

Real life library			
Diagnoses	Number of images (n=283)	Percentage	% Textbook library
Basal cell carcinoma	43	15%	13%
Squamous cell carcinoma	8	3%	8%
Melanoma/Melanoma in situ	9	3%	18%
Viral wart	9	3%	4%
Cysts - benign	4	1%	2%
Melanocytic naevus- benign	85	30%	19%
Seborrheic keratosis	69	24%	6%
Actinic keratosis	18	6%	3%
Bowen's disease	25	9%	3%
Dermatofibroma	7	2%	4%
Keratoacanthoma	1	<1%	6%
Lipomas	0	0	2%
Pyogenic granuloma	5	2%	3%
Mycoses fungoides	0	0	6%
Pagets disease	0	0	3%
Cutaneous metastases	0	0	0%

Table 8.3c: Table showing the diagnostic subcategories of the real-life library images.

Whilst when the percentages of the diagnostic subcategories are compared, many are similar (see the two righthand side columns of Figure 8.3c), hardly surprisingly there are striking differences in the over-representation of melanoma and the under-representation of seborrheic keratoses in the textbooks. As in the sequentially photographed real life library only nine melanomas presented during the 9-month

period the decision was made to enrich this category with randomly selected melanoma images from the Dermofit library. In consequence, 20% (13/65) of the melanoma images available in the Dermofit library at the commencement of the study were randomly selected and added to the real-life image text set.

Study 1

The students' diagnostic accuracy was assessed in a test slideshow of 60 randomly selected pigmented lesions. Three pigmented diagnoses were chosen to include melanomas and its most commonly confused differentials (seborrheic keratosis and benign melanocytic naevi). All three lesional classes were well represented in the textbook library.

The 60-image slideshow was constructed in a randomly stratified manner with 30 images from the textbook library and 30 images from the real-life library. Within each set of 30 images were 10 lesions for each of the three pigmented diagnoses examined; benign melanocytic naevi, melanomas and seborrheic keratoses. All the images selected (both the scanned "textbook" images and the digitally photographed "real-life" images) were cropped to contain a single lesion. The resultant images were inserted into a Keynote (APPLE, California) slideshow with a white surround. In Keynote they were centred and resized to the same proportions (195cm²). The order of the 60 slides was then randomly determined.

The test slideshows were performed using a 15" Macbook Pro (APPLE, California) connected via a HDMI link to a LG 55" HD TV (LG, South Korea). Each slide was displayed for 20 seconds and students recorded their answers on purpose designed answer sheets. Again, marking was "generous" in that answers were accepted as correct despite spelling mistakes, incomplete terminology, abbreviations and lay terms.

To facilitate the analysis of the results and allow the exclusion of confounding factors all students recorded their use of supplementary learning materials (textbooks, websites etc.) on the reverse of their answer sheets. To minimise the possibility of

fatigue effect the image order in the slideshows was randomly reallocated for each student cohort. To limit any effect of selection bias a second randomly stratified 60-image slideshow batch was constructed in the same manner as described above for the final two student cohorts. The melanoma and melanocytic naevi images from both slideshows are displayed below in Figure 8.3d.

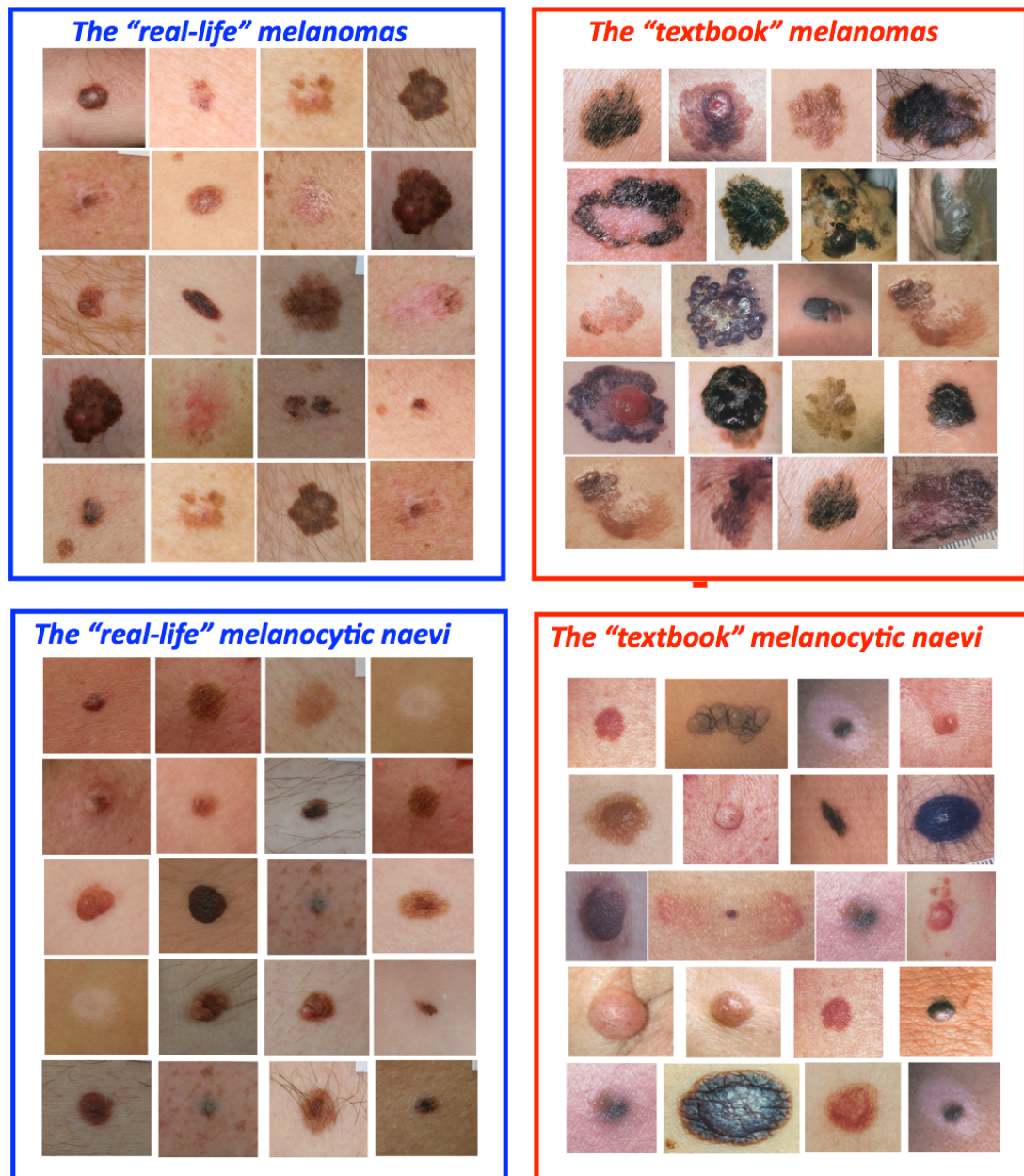


Figure 8.3d: The melanoma and melanocytic naevi images randomly selected from the "real-life" and "textbook" libraries for Study 1.

As a proxy measure as to whether there were inherent differences between the images of the real-life and textbook libraries that could influence the results, five dermatology registrars (three female, two male) were asked to sit in on both batches of the student test slideshows and assess each of the 60 images as to whether they thought they came from a textbook or a clinical photograph.

Study 2

The students' and registrars' diagnostic accuracy, speed and confidence was assessed in an individual undertaken test slideshow of 40 skin lesions. The 40 lesion test slide show was randomly stratified to contain 20 'real-life' images and 20 'textbook' images. Within each set of 20 images there were two melanomas; two melanoma-in-situ; four SCCs; four BCCs; four seborrheic keratosis and four benign naevi. The computer display and set up was as per Study 1 except that as this was an individual study there was no defined time limit to the subject's exposure to the test image. The subjects progressed to an intervening blank screen once they had offered a diagnosis for the lesion image. A handheld stopwatch was used to record the time between the image first appearing and progression to the blank screen (to the nearest 0.1s). Subject's diagnostic confidence for each diagnosis was also noted, according to a 7-point Likert like scale (1=no confidence; 7=absolute confidence).

8.4 Results

Study 1

Sixty-one students completed the test slideshow, representing 98% (61/62) of students asked to participate. The students' overall accuracy was 35% (1298/3660). There was no difference in overall accuracy between the sexes. There was also no difference in accuracy between the four different student cohorts or between the two randomly stratified batches of 60-images. Wilcoxon paired test showed that the students' overall accuracy was not significantly different between the two image libraries; with a median score of 11 lesions correctly identified and accuracy of 37% (669/1830) for the textbook images, compared to a median score of 10 lesions identified and accuracy of 34% (629/1830) for the real-life images- see Figure 8.4a. However for the individual classes of lesions significant differences were identified.

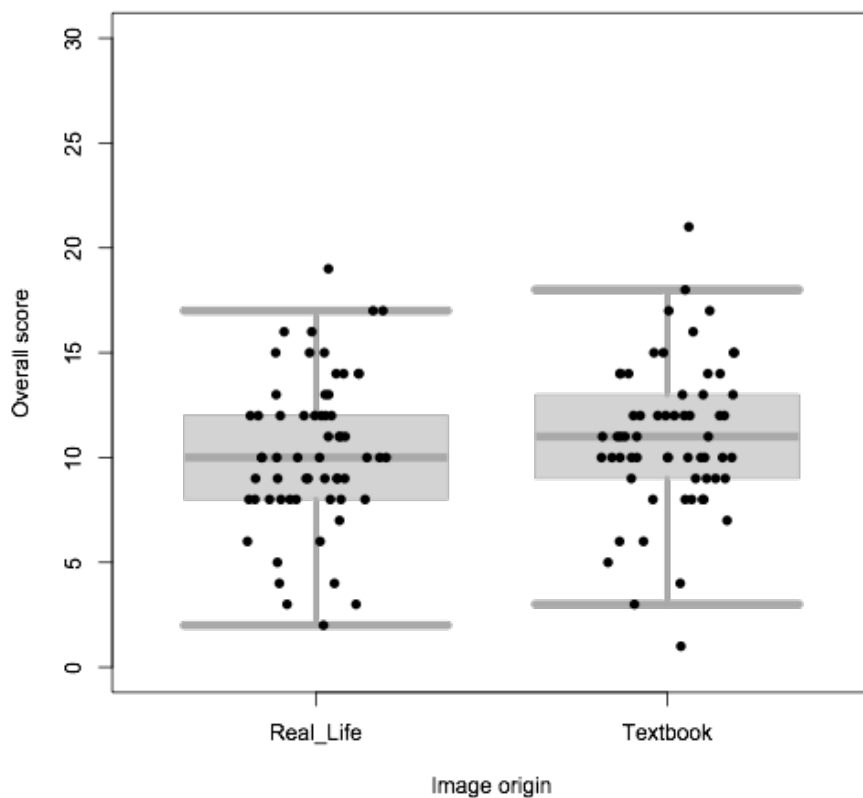


Figure 8.4a: Combined scatter and box plot demonstrating the subjects' overall scores by image origin.

Melanoma images

Across the study the median number of correctly diagnosed real-life melanomas was 4 with an accuracy of 37% (227/610) in comparison to the textbook melanomas where a median of 7 were identified correctly with an accuracy of 68% (412/610). See Figure 8.4b. A Wilcoxon paired test confirmed these differences to be significant ($p < 0.001$). These differences were seen in each of the four student cohorts enrolled ($p < 0.001$). 4% (24/610) of the textbook melanomas were under diagnosed as moles in comparison to 19% (116/610) of the real-life images.

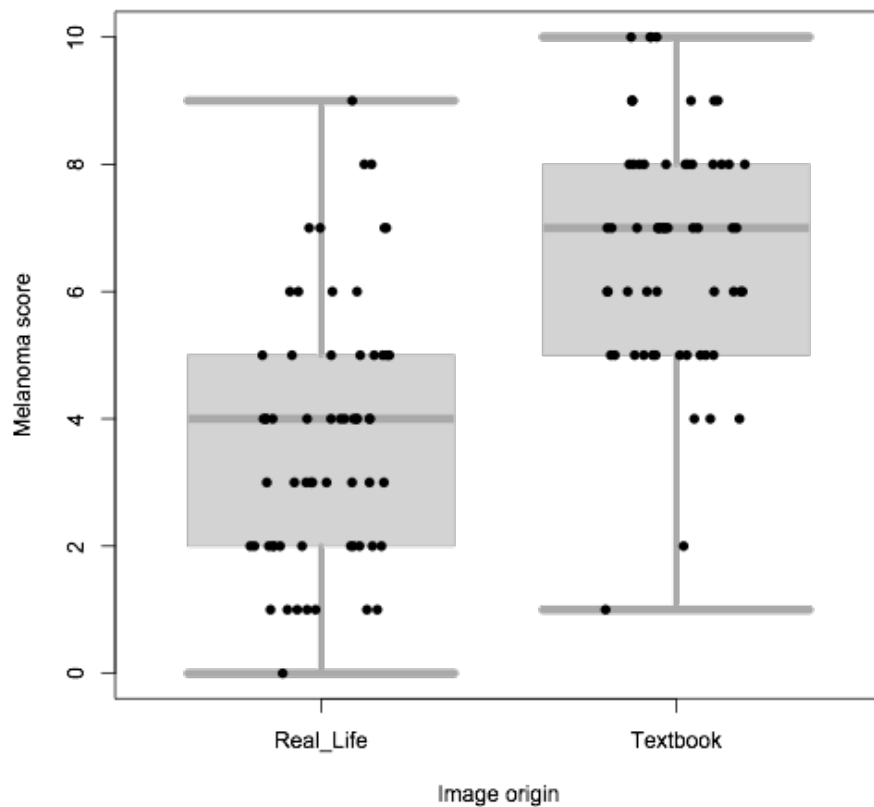


Figure 8.4b: Combined scatter and box plot demonstrating the subjects' melanoma scores by image origin (real-life vs textbook).

Melanocytic naevi images

For the images of melanocytic naevi the median number correctly identified was 4 with accuracy of 40% (243/610) for the real-life images, compared with a median of 2 with an accuracy of 21% (129/610) for the textbook images. A Wilcoxon paired test confirmed these differences to be significant ($p < 0.001$). Again these differences were seen in all four of the student cohorts enrolled ($p < 0.001$). 31% (190/610) of the textbook melanocytic naevi were over diagnosed as melanomas in comparison to 19% (116/610) of the real-life melanocytic naevi.

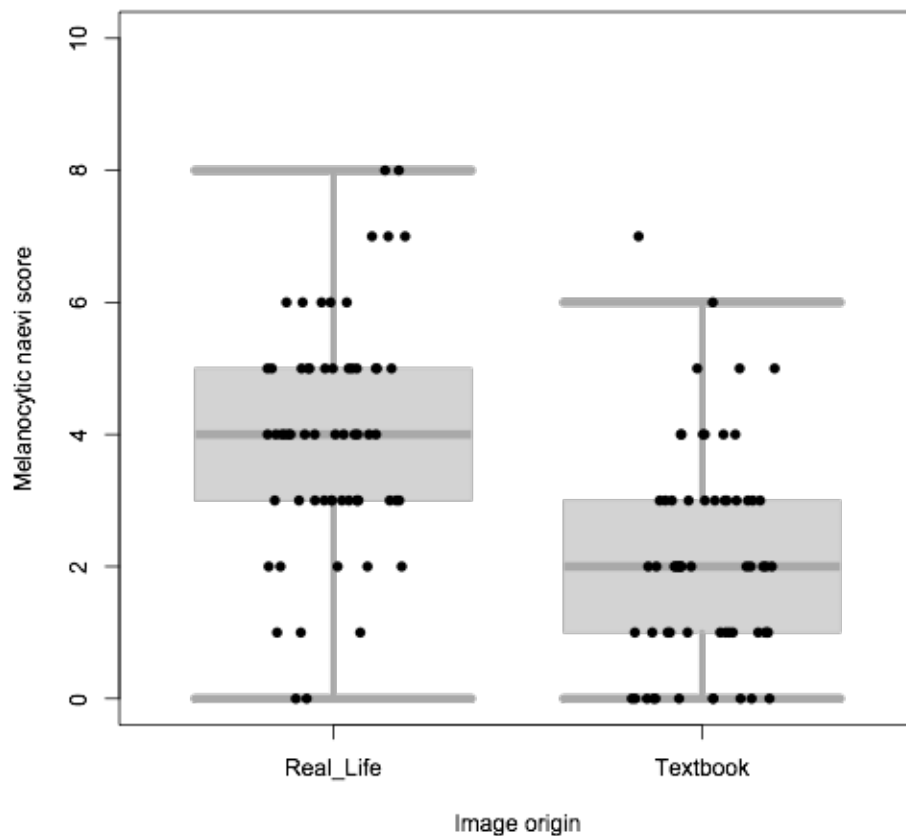


Figure 8.4c: Combined scatter and box plot demonstrating the subjects' melanocytic naevi scores by image origin (real-life vs textbook).

Seborrheic keratoses images

In contrast, there were no significant differences observed between real-life and textbook libraries for the seborrheic keratoses images. The median number of seborrheic keratoses identified correctly was 3 with an accuracy of 26% (159/610) for the real-life images compared to a median of 2 and accuracy of 21% (128/610) for the textbook images.

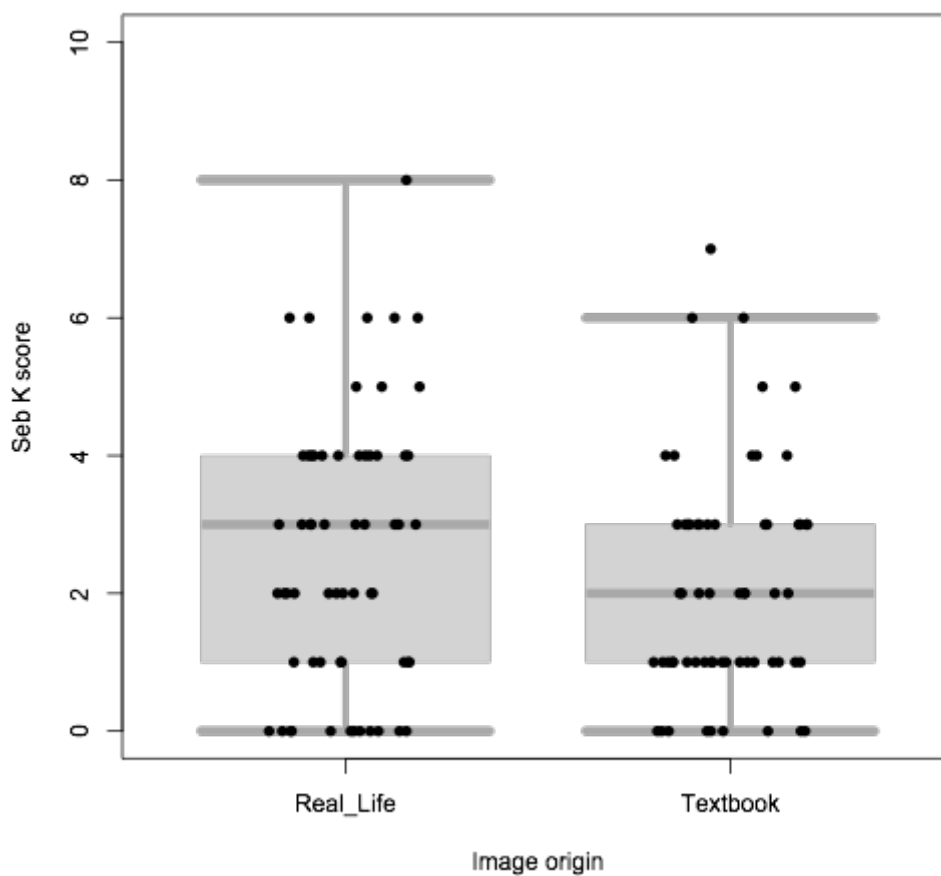


Figure 8.4d: Combined scatter and box plot demonstrating the subjects' seborrheic keratosis scores by image origin (real-life vs textbook).

Registrar proxy control

The five dermatology dermatology registrars had an overall accuracy for predicting the origin of the images of 54% (326/600). There was no difference between the sexes, nor was there any difference between their performance in the two slideshow batches.

Students' previous exposure to the textbook images

Based on the students' self-reported use of textbooks which they recorded on completion of the test, it can be calculated that of the 3660 images tested in the study a possible 2% (71/3660) could have been viewed prior to the test. Whilst it was logistically impossible to exclude these from the analysis it was possible to demonstrate that the students' diagnostic accuracy for these 71 images did not differ significantly from the other 1759 textbook images tested. Therefore, the study findings are unlikely to have been affected by the minor degree of previous exposure.

Study 2

The fifteen students enrolled in Study 2 correctly identified a median of 9 lesions with an accuracy of 48% (144/300) for the textbook images compared to identifying a median of 7 lesions with an accuracy of 30% (91/300) for the real-life images. see Figure 8.4e. Wilcoxon paired test confirmed these differences to be significant ($p < 0.01$).

The eight dermatology registrars showed the same pattern; with a median of 14.5 lesions identified correctly from the textbook images and an overall accuracy of 73% (117/160) compared to a median of 10.5 lesions and an overall accuracy of 53% (84/160) for the real-life images. Again, this difference was significant (Wilcoxon paired test $p < 0.01$).

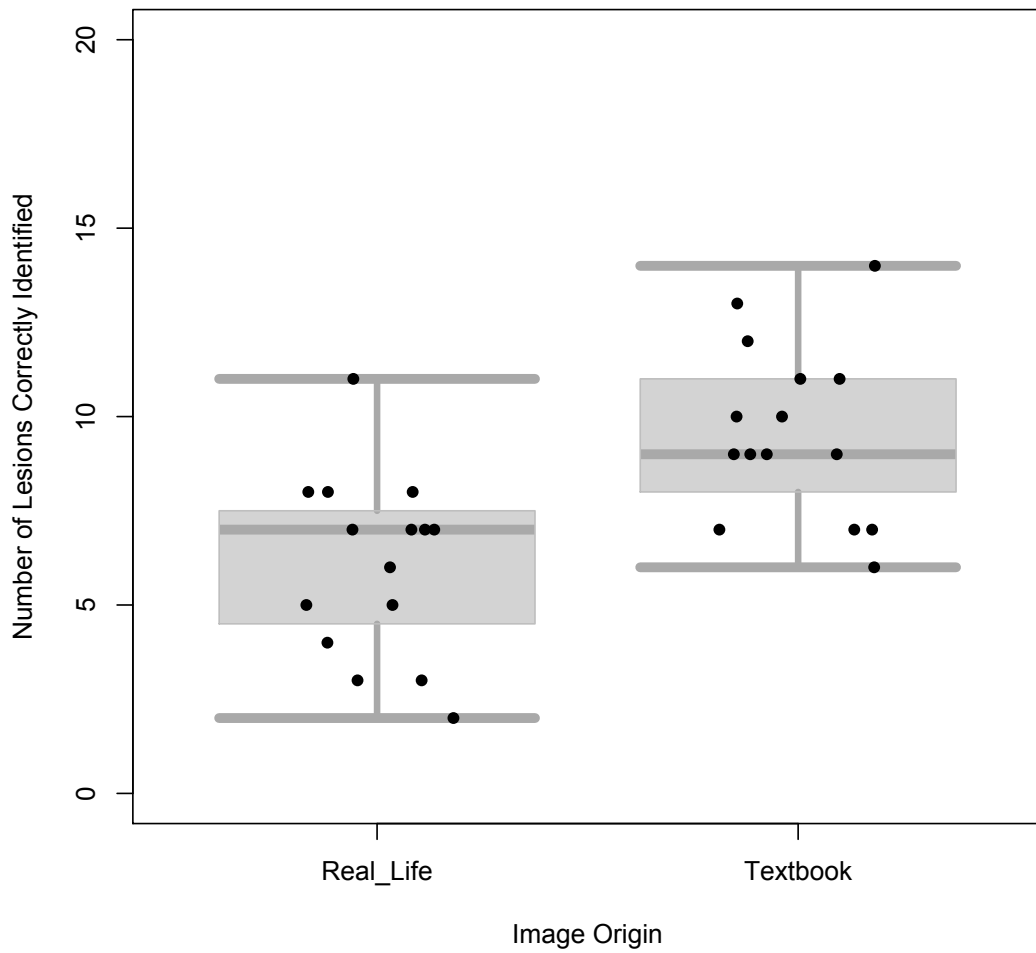


Figure 8.4e: Combined scatter and box plot demonstrating the students' overall scores by image origin (real-life vs textbook) for Study 2.

The trend of improved accuracy in the textbook images was seen across all the lesion classes and both cohorts, except for the diagnostic category of textbook melanocytic naevi, where the finding of Study 1 were replicated. The students assigned higher Likert confidence ratings to the textbook images with a median Likert score of 5 compared to a median Likert score of 4 for the real-life images (Wilcoxon paired test $p < 0.01$).

Timings

Overall the students diagnosed the textbook images quicker than the real-life images with means of 6.4 secs and 7.6secs respectively (Paired t-test $p < 0.01$). The same

findings were replicated in the dermatology registrars at 3.5sec vs 5.3secs (Paired t-test $p < 0.01$). The students' mean time for correct diagnoses was 4.4secs compared to 5.6secs for incorrect diagnoses (Paired t-test $p < 0.01$). Again, this was replicated in the dermatology registrar cohort at 2.9secs for correct diagnoses and 3.9secs for the incorrect diagnoses (Paired t-test $p < 0.01$). The time taken to diagnose a lesion (whether correctly or incorrectly) was inversely related to the students' diagnostic confidence - see Figure 8.4f.

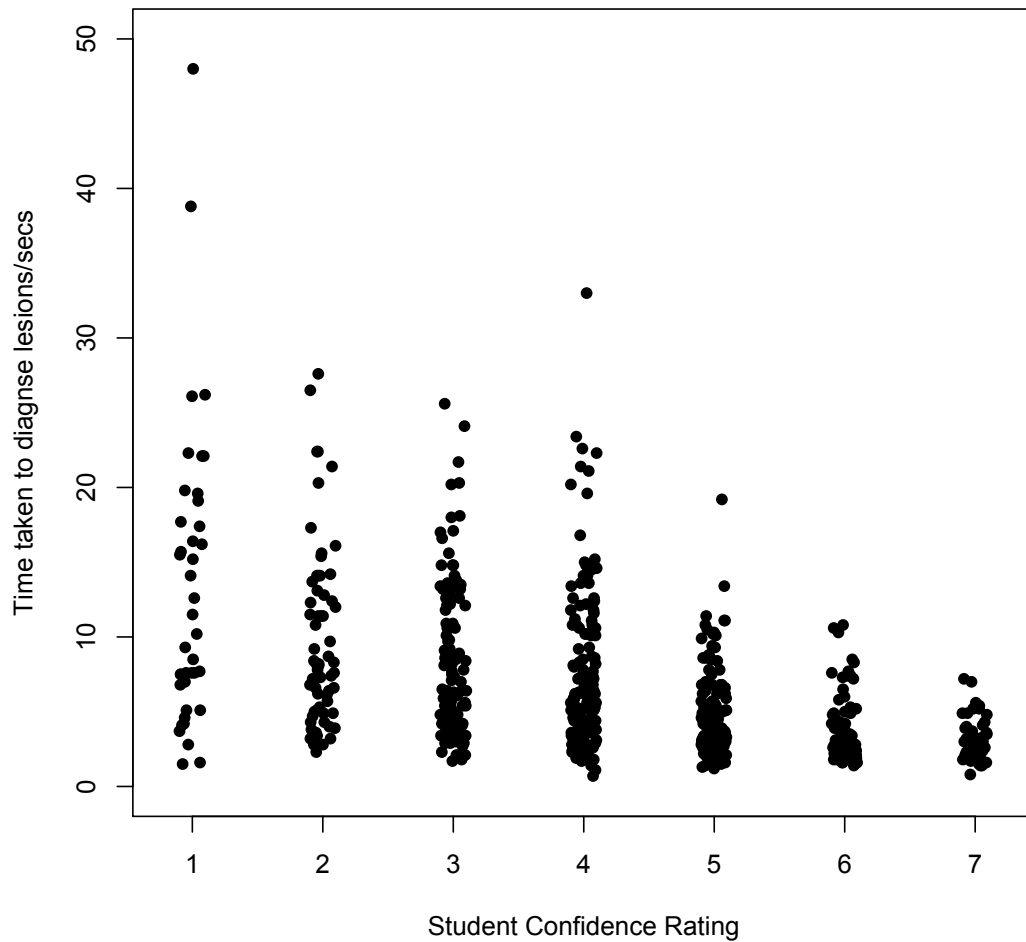


Figure 8.4f: Scatter plot showing the time students took to diagnose a lesion against their Likert confidence ratings for that lesion (1=no confidence; 7=absolute confidence).

8.5 Discussion

As the dermatology registrars were unable to identify the source of the images, and hence whether they were from the textbooks or the clinic, it can be assumed that any differences observed between the subjects' diagnostic accuracy between the two image datasets was unlikely to be related to the quality of the images.

Although it was not possible to entirely control for the students' previous exposure to the textbook images, less than 4% of the textbook image answers could have been affected by prior exposure. Analysis also revealed no significant difference in the students' accuracy in diagnosing these images compared to those in textbooks they had not seen.

Students clearly found textbook melanoma images significantly easier to diagnose. The reverse was observed for melanocytic naevi. It is apparent from a perusal of the images (see Figure 8.3d) that the textbooks contain more examples of advanced melanoma and more atypical examples of melanocytic naevi than are ordinarily encountered in real-life practice. If such illustrations are the primary source from which diagnostic insight is gleaned it can be readily appreciated why (because of a possible confusion with atypical naevi) in Study 1 the subjects exhibited a tendency to under diagnose early real-life melanomas.

Study 2 demonstrates that when tested on a bigger range of diagnostic subcategories (including NMSC as well as the pigmented lesion groups in Study 1) there was a significant difference in overall diagnostic accuracy. The subjects found the textbook images significantly easier to diagnose and were far more confident in making these diagnoses. This increased confidence correlated to an increased speed of diagnosis, lending support to NAPR being the dominant mechanism underlying their diagnoses. This is in keeping with other studies, where speed of object identification has been suggested to correlate with NAPR and higher diagnostic accuracy (212-214).

More worryingly this study lends support to the impression that the development of students' NAPR may be distorted by exposure to selected "classical" images, in that

their confidence in dismissing the possibility of melanoma when a melanocytic lesion exhibited only subtle changes was worryingly high. This study therefore reinforces work presented elsewhere in this thesis suggesting that exposure to a larger number and a wider variety of images is likely to lead to improved diagnostic acumen. Such findings also have implications for non-expert skin cancer education and the subsequent ability of such non-experts to accurately identify skin cancers; but also emphasises the importance of multiple images may play in any public image based skin cancer education strategies.

Chapter 9

Exteriorising the experts' internal reference library for skin lesions

9.1 Introduction

The work presented in this chapter has been published in Acta Derm Venereol. 2011; 91(3): 279-83. PMID: 21461552. "Utility of non-rule-based visual matching as a strategy to allow novices to achieve skin lesion diagnosis."

The previous chapters have highlighted the role of non-analytical pattern recognition (NAPR) in the identification of skin lesions and have shown that novices have intrinsic ability to appreciate similarity between lesions with a high degree of consistency. Furthermore, work presented has shown that exposure to multiple examples facilitates the acquisition of diagnostic accuracy. These findings accord with both Schmidt's and Norman's suggestion that experts develop internal exemplars of disease (55,62,63,140), but the work presented also suggests achieving such expertise would require more commitment of time than is practicable for a non-expert.

The study presented in this chapter investigates the possibility of exteriorizing, to some degree, the experts' library of exemplars. If untrained novices can reliably distinguish diagnostic categories of skin lesions based on how they look, irrespective of their previous exposure, it should be possible to build diagnostic support tools that can enhance their natural pattern-recognition abilities whilst compensating for their lack of experience. A "gedanken experiment" elucidates the role of such a diagnostic support tool.

If an individual presents with a skin lesion and a non-expert has access to an infinite photographic library of skin lesions, in theory (provided the non-expert can appreciate the relative visual similarity or dissimilarity between different lesions) a visual match should be achievable between the index lesion and a diagnostically tagged referent image: and thus a diagnosis made. The major obstacles to such an approach are the

absence of an infinite library and the time taken to successively match the index case with a series of referent images. The time element can be reduced if it is possible to order the image referent library in much the same way as a physical library is indexed by subject, author and so forth. The users of such a tool are (by definition) inexperienced, so the ordering system must be intuitive as current skin atlas cataloguing methods which require a posteriori knowledge are unsuitable.

Is it therefore possible to order images of skin lesions using some attributes of visual 'likeness' which would be perceived as common by different people? Furthermore, is the trade off in accuracy between an infinite library and one that is merely large acceptable in that are novices able to match not merely identical images but ones that closely related?

9.2 Aim

To develop a computer software system that would allow the user to make a direct visual comparison of an index image to a diagnostically tagged reference image library and to test such a systems utility using novices and medical students.

9.3 Materials & Methods

Subjects

Study 1:

All students that attended for their two-week clinical attachment over a 3-month period (November 2009- January 2010) were recruited into the study. In total 60 students were enrolled (4 batches of between 14-16 students). Thirty-six (60%) of the students were female.

Study 2:

20 lay members of the public were recruited between May and July 2010. Mean age was 33 (Range 21-61). 75% were female. All but four had completed university education and the 20 subjects were employed in a wide range of different occupations (eg solicitor, accountant, teacher, secretary, chef). No volunteer had any personal experience of skin cancer nor had undergone any tuition in the identification of skin lesions.

Software Image Selection

Eighty images from five diagnostic classes of commonly referred focal skin lesions were selected from the Dermofit image library. The images comprised 14 haemangiomas, 23 seborrheic keratoses, 19 melanocytic naevi, 15 basal cell carcinomas and 9 squamous cell carcinomas. Images were chosen based on technical quality of the images and because they were thought to be representative of a specific visual diagnostic class. The images were determined by the author printing out thumbnails of the entire validated Dermofit library, as it stood at the start of this study and sorting the images into what were, in his opinion, similar visual diagnostic subclasses i.e. a larger version of the free sorting described in Chapter 3's Study 1.

These five diagnostic groups make up the majority of the lesions that are referred from primary care for specialist assessment. From these 80 images, twelve index lesions were randomly selected, with the remaining 68 images acting as referent images in the software image database.

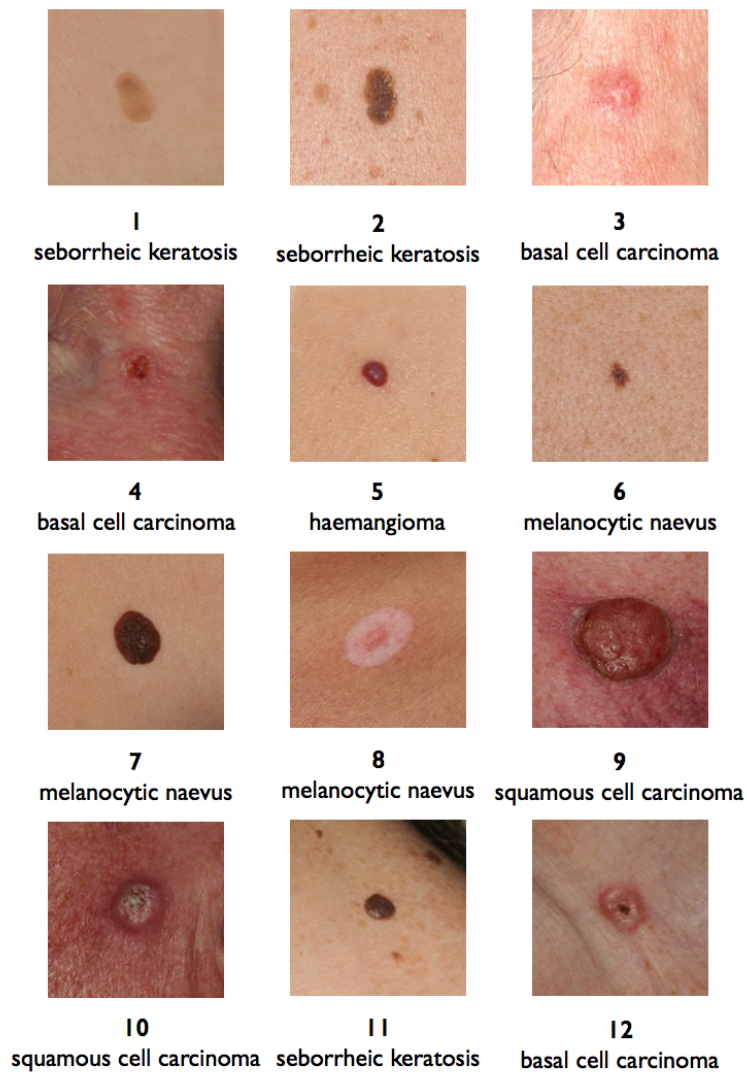


Figure 9.3a: The twelve index images randomly selected from the 80 lesions selected for the prototype software.

Software Design

The prototype software allows the user to make a direct visual comparison between a centralised index image and up to twelve surrounding referent images. The user then navigates through the library of referent images until they feel that they have successfully matched the index lesion to a similar referent image (or images).

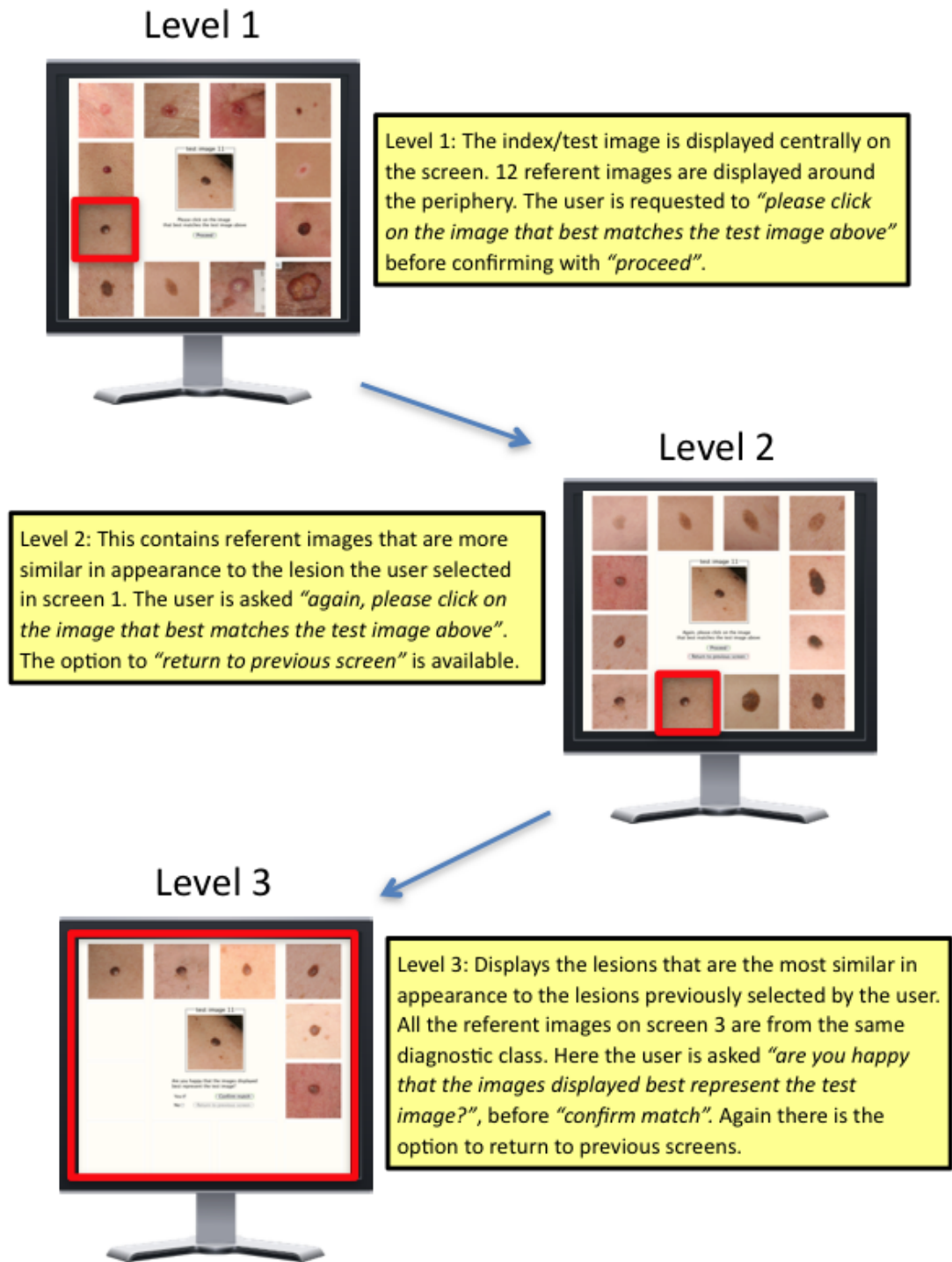


Figure 9.3b: Schematic showing the three level setup of the software developed - with real screen shots from the software integrated into the diagrams.

In this experiment, the 68 referent images were arranged over three levels utilising a total of 18 different screens (one screen for level 1 with the initial 12 referent images,

five screens for level 2 comprising each of the five diagnostic classes, twelve screens for level 3 for each of the author's twelve visual subclasses). Irrespective of which index image was being tested, the referent images in the first level's screen were identical for all matching attempts and consisted of an image from each of the visual subclasses. It was only the subsequent second and third level screens' referent images that were determined by the individual user's image selection. The order in which these five second level and twelve third level screens were displayed and their relationship to a specific users image selection was predetermined by the investigator and was kept constant for the duration of the experiment. The method employed for grouping the 68 images to the 18 screens and the relationship of a screen to a specific user interaction was based on the author's opinion of visual similarity and the lesions' underlying pathological diagnosis and assigned visual subclass. If the user was not content with their selection at any stage of the process (prior to confirming their final match) the software allowed them to retrace their steps. The software was written in JAVA in conjunction with the Department of Informatics and was run on 20" iMACs (APPLE, California). Although the screen-snap shots attest to the software being moderately intuitive (see Figure 9.3a), nonetheless to demonstrate how to navigate through the software library and how to make a final diagnostic match a short instructional video was integrated within the software. To avoid any potential bias, the video did not include images of skin lesions but demonstrated the key features of the software using simple pictures of differing shapes (circles, squares, crosses).

Study 1

On the morning of Day 1 of the dermatology attachment (prior to seeing any patients), each batch of students was randomly split into two groups; the first group (in total 31 students) was asked to identify each of the twelve index images using the software and the second group (in total 29 students) was asked to identify the twelve test images by writing their diagnosis on an answer sheet. Forthwith the students using our prototype software to obtain a diagnostic match will be addressed as the 'software' group, and the students that provided written answers as the 'control' group. Test instructions were standardised across the batches of students. As in previous studies, answers were "generously" accepted as correct for the control group; allowing spelling

mistakes, incomplete terminology, abbreviations and lay terms. After the Day 1 test no score or feedback was provided to either group. Exactly the same experiment was repeated on the afternoon of Day 10 at the end of the students' dermatology attachment. The format of both the Day 1 and Day 10 experiments was identical, except that the introductory software video was not repeated to the software group on Day 10.

The twelve test images were presented to both groups of subjects in the same order and in an identical format. For both the software and control groups the skin lesion images were displayed using the same 20" iMac monitors (APPLE, California) with identical resolutions (1650x1050), calibrated for colour inconsistencies using the Pantone Huey Pro calibration (PANTONE LLC, New Jersey). The experiments were all undertaken in a designated curtained room with similar ambient lighting conditions. No time restrictions were imposed for either group. Constructive feedback was only given after each batch of students had completed the Day 10 test during an additional tutorial.

Study 2

The 20 subjects were provided with the same introductory video guide to the software as the students but no additional training. The experimental setup was identical to that undertaken by the students, with the same twelve test images and an identical version of the software (as described above). This group of subjects will subsequently be referred to as the 'lay' group.

9.4 Results

Study 1

93% (112/120) of students completed both the Day 1 and Day 10 test. Student absence was evenly distributed across the four test groups; Day 1 control group (n=1), Day 10 control group (n=3), Day 1 software group (n=1), Day 10 software group (n=3). At the start of their dermatology attachment (Day 1 test), out of the twelve test

images, the control group correctly diagnosed a median of 1 image with a diagnostic accuracy of 16% (55/336), in the same Day 1 test the software group correctly identified a median of 12 images, resulting in a diagnostic accuracy of 99% (357/360). At the end of the students' dermatology attachment (Day 10 test) the control group correctly diagnosed a median of 6 images with a diagnostic accuracy of 51% (160/312) and the software group matched 12 images correctly, with a diagnostic accuracy of 99% (335/336).

Two sample Wilcoxon tests showed that the scores at Day 1 between the software and control group were significantly different ($p < 0.0001$) as were the two groups scores at the end of the students' attachment on Day 10 ($p < 0.0001$). Wilcoxon paired test showed that the control group's scores improved significantly ($p < 0.0001$) over their attachment, whereas the software groups score did not appear to change ($p = 0.582$). There was no difference in test scores between the four batches of students or between the sexes. In addition, no particular pattern of results was identified with respect to lesion type. The results for all the subject groups enrolled in both Study 1 and 2 are plotted below – see Figure 9.4a.

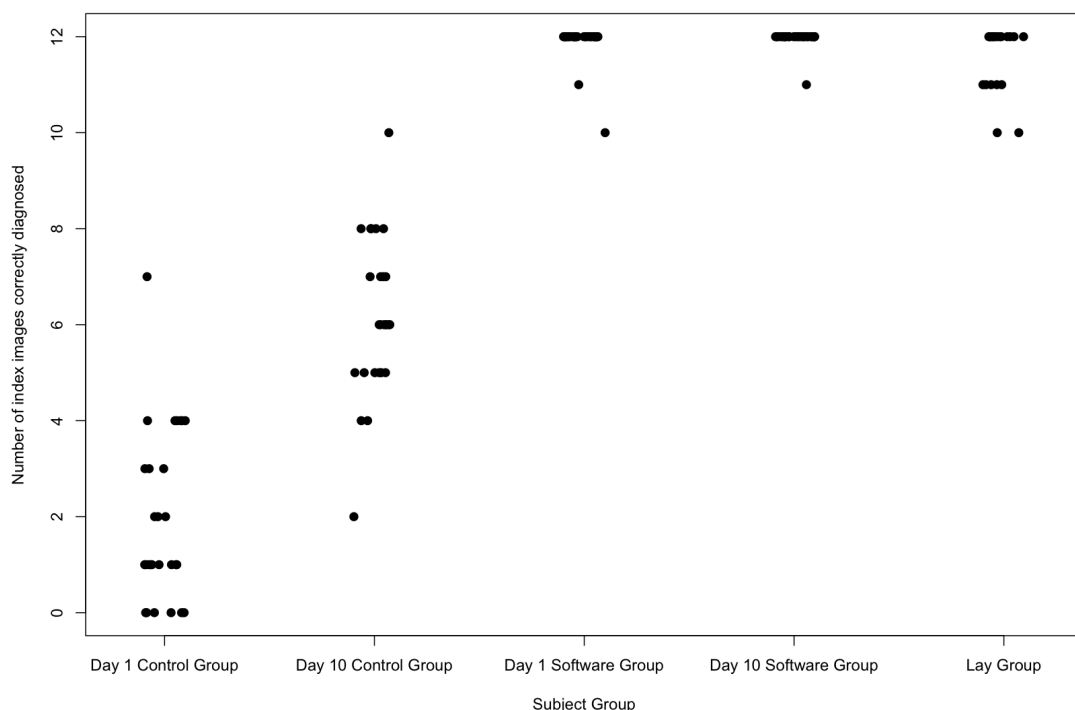


Figure 9.4a: Plot showing all the subjects' scores enrolled in Studies 1 and 2.

Study 2

The lay group, using the software, correctly identified a median of 12 images resulting in a diagnostic accuracy of 96% (231/240). Again, there was no difference in test scores between the sexes or with respect to lesion type. Two sample Wilcoxon tests showed that the student control group had significantly inferior diagnostic accuracy when compared to the lay group, at both the start and end of their dermatology attachment ($p < 0.0001$). There were no significant differences in overall test scores between the sexes or with respect to the diagnostic class or specific test image.

Subclass accuracy

On the same twelve test images, further analysis shows that the novices correctly identified the visual subclass (as defined by the author) in a median of 11 images (Range 9-12) with a total subclass accuracy of 92% (221/240). By definition the subclass accuracy can only be lower than the overall diagnostic accuracy as not only is it a more complex and subjective task, but fundamentally the software structure made it impossible to get the subclass correct but the overall diagnosis wrong. This was due to its inherent structure (see Figure 9.3b above and its' accompanying text) where the subclass matching was achieved on the final level 3 screens after the diagnostic class match had been achieved correctly by the participants on a level 2 screen. Like the overall diagnostic accuracy the subclass accuracy did not vary by sex or by diagnostic class, however it did significantly vary according to the test image (Chi-squared $p < 0.001$) - see Table 9.4b below.

Test Lesion	Diagnostic Class	Overall Class Accuracy	Visual Subclass	Subclass Accuracy
1	Seborrheic Keratosis	100% (20/20)	Seb K 1	100% (20/20)
2	Seborrheic Keratosis	100% (20/20)	Seb K 2	100% (20/20)
3	Basal Cell Carcinoma	100% (20/20)	BCC 1	95% (19/20)
4	Basal Cell Carcinoma	100% (20/20)	BCC 2	90% (18/20)
5	Haemangioma	95% (19/20)	Vasc	90% (18/20)
6	Melanocytic Naevus	100% (20/20)	Naevus 1	100% (20/20)
7	Melanocytic Naevus	90% (18/20)	Naevus 2	90% (18/20)
8	Melanocytic Naevus	100% (20/20)	Naevus 3	100% (20/20)
9	Squamous Cell Carcinoma	95% (19/20)	SCC 1	95% (19/20)
10	Squamous Cell Carcinoma	80% (16/20)	SCC 2	80% (16/20)
11	Seborrheic Keratosis	95% (19/20)	Seb K 3	95% (19/20)
12	Basal Cell Carcinoma	100% (20/20)	BCC 3	65% (13/20)
Full Test	NA	96% (231/240)	NA	92% (221/240)

Table 9.4b: Table detailing the novices' subgroup accuracy for the twelve images test images.

The lesion 'BCC 3' stands out with a subclass accuracy of only 13/20 (65%), despite the fact the overall accuracy for all of the three BCC test images was 60/60 (100%). The fundamental role of a diagnostic support tool is facilitating a high diagnostic accuracy; although the prototype tool achieved this with 100% diagnostic match for the test image 'BCC 3', the lesion's low subclass accuracy demonstrates the sometimes subjective nature of subclassing but, as is discussed below, such 'errors' in subclass accuracy are not necessarily a weakness of the approach, rather they can be used to redefine the subclasses and improve the cataloguing process.

9.5 Discussion

The results show clearly that within the constraints of a limited range of diagnostic possibilities using an image based approach, that medical students can utilise visual matching as a diagnostic strategy and achieve diagnostic scores higher than medical students who have completed a standard clinical dermatological attachment. The ability is not confined to medical students as a group of lay novices scored similarly. The success was therefore achieved by participants matching images based on visual similarity without any attempt to apply explicit rules of likeness based on dermatological knowledge. These results support the concept that it might be possible to exteriorise an experts' library of exemplars to enable novices to achieve diagnosis by NAPR.

Immediately after completing their undergraduate dermatology teaching attachment, students' unaided diagnostic accuracy of common skin lesions was 51%. Although this reflected an improvement in their baseline diagnostic accuracy over the course of their two-week attachment and accords with an improvement in final diagnostic accuracy to 33% (on a wider range of lesions) in the study presented in Chapter 3. However disappointing these students' scores may seem, they are, in fact, not dissimilar to previous studies which have investigated the diagnostic accuracy of non-dermatologists after medical school training with colour images (187,215).

Although it may not be possible to directly extrapolate these results to clinical practice, so low a final diagnostic accuracy is of concern. Albeit any absolute score is dependent on the difficulty of the test set and in practice when diagnosing easier lesions overall accuracy may improve. However, the results are more sobering if one considers that the level of these students' diagnostic accuracy may reflect an artificially raised result; as the students in the study presented in this chapter achieved their level of accuracy after double-exposure to the twelve test images (the students had previously viewed, albeit without feedback, the twelve images during the first test on Day 1 of the attachment). Although the result of this study underlines the need to define ways in which undergraduate teaching of dermatology can be improved, if expertise is dependent on acquiring a personal databank of reference images there

may be an upper limit to the diagnostic expertise achievable by a non-expert. This may be overcome either by recruiting a cadre of non-medical lesion assessment experts or ensuring future lesion databases are freely available online to non-expert medical practitioners such as GPs who are responsible for the primary screening of potentially malignant lesions.

The development of the prototype software described in this study is indeed intended to make that expert database available to the non-expert but several problems will have to be surmounted before any large-scale adoption of this software technique allows exteriorisation of the experts' library to assist in the diagnosis of skin lesions.

First, the testing relied on matching to an image not to a lesion on a real patient. Although there remains some uncertainty as to whether it is possible to directly extrapolate image-to-image matching to patient-to-image matching, images are widely used in teaching and in the examination of clinical competence along with teledermatology applications. It has been demonstrated elsewhere in this thesis that this approach can be successfully used for teaching, so it would seem reasonable to deduce that if it is possible to match virtual patients it should be possible to match real patients. If such a matching tool is envisioned as a diagnostic support tool for non-experts in medical practice or for the lay public (for instance in encouraging early presentation of suspicious pigmented lesions) this potential limitation needs further exploration.

Second, in the present studies there was no attempt to represent the whole of the complexity of dermatological morphology, as it focused on a small range of common lesions. Although ultimate performance figures must relate to the difficulty or atypicality of the match, this study is a positive proof-of-concept suggesting further work is merited.

The study was based on only 80 images. It is, however, eminently scalable provided software is written that will allow matching of several thousand images. It is likely that as the software database increases in size it will become increasingly powerful provided it can be ordered in a manner that is intuitive to the user. Unlike this

software prototype, where the initial selection and ordering of the images was undertaken visually by the author, large scale ordering of images will need to be based on automatically extracted properties ('computer vision'), user feedback, or some combination of the two. The work allied to these concepts was the subject of the Dermofit projects collaborations with the Department of Informatics and contained within the Appendix of published work.

That novices could identify skin lesions without any explicit definition of likeness or specific rule based analysis (such as the ABCD) makes this approach fundamentally different to most previous strategies designed to improve non-expert diagnosis. Whilst it is tempting to explore exactly what features of the referent images the users are unconsciously matching, this may neither be necessary nor tractable and the implications of this in regard to the fully automated assessment of dermatological lesions will be discussed later in the thesis.

Chapter 10

Hazards of single lesion assessment

10.1 Introduction

The work presented in this chapter has been published in Acta Derm Venereol. 2013; 93(6): 689-92. PMID 23695107. "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."

Traditional clinical teaching emphasises the importance of a full clinical examination in accurate patient assessment and the initiation of appropriate clinical management. In the assessment of potentially neoplastic index skin lesions there exists several cogent arguments for examining the whole of the patient's skin (i.e. a total body skin exam, hereafter abbreviated to TBSE).

First, the incidence rate of skin cancer is such that not unexpectedly, as the whole skin has been subjected to similar host and environmental factors, patients often present with more than one cancer at a single time point. The rates of synchronous and metachronous skin tumours cannot be readily calculated from the national Scottish data due to the nature of the cancer registration process, but other regions (albeit with higher baseline incidence rates) suggest the problem is considerable with rates of 20% and 20-100% respectively (216-220). Clearly as any one person cannot readily visualise their whole integument, individuals may be unaware of any such additional lesions.

Second, the diagnostic process for any particular lesion is often influenced by factors beyond the index lesion, such as background solar induced skin damage, but perhaps the clearest example is the 'Ugly Duckling' sign, in which the assessment of an individual pigmented lesion is heavily influenced by the morphology of the other non-index pigmented lesion (60). This example of intra-patient differential pattern recognition has been shown to be highly reliable amongst dermatologists and an

essential component of efficient screening for melanomas (56,221,222).

Although little systematic data on this topic has been identified, the author has been increasingly aware of a tendency for clinicians – both in primary and secondary care – not to fully undress patients who present with index lesions for which skin cancer is a possible diagnosis or during clinical follow up for previously diagnosed tumours. The reasons for this variation from “gold-standard” practice are multiple, but probably relate to time pressures in clinical practice, changing norms concerning patient consent and chaperoning, all of which involve additional costs that add further pressures in today’s cost sensitive health service. Whatever the cause, the tendency to examine only a single index lesion has perhaps generated scepticism of the necessity and importance of searching for incidental lesions and hence the value of a TBSE.

This trend to examine only the presenting index lesion is further reinforced by the ongoing pressures to replace conventional clinical examination with teledermatology. The term “teledermatology” was first coined by Perednia and Brown in 1995 (223). In this paper, the authors suggested that the dermatology services of rural communities might be improved if digital images of skin conditions were transmitted to metropolitan dermatologists to allow them to undertake remote consultations. The authors highlighted that in large areas of the USA, especially in mid-west states, access to dermatologists was inadequate. In their home state of Oregon, which has a landmass approximately 10% greater than UK, only two dermatologists covered the eastern two-thirds of the state.

Whilst the initial publications that followed continued to explore teledermatology’s utility in serving rural populations (224), within a few years papers started to suggest that teledermatology may have potential for triaging non-rural referrals (225). Initially the proponents of teledermatology triage suggested that such an approach could prioritise frankly malignant referrals thus expediting their treatment, but over the last few years there has been an increasing focus towards the cost savings that may be achieved by excluding the benign referral lesions from the traditional process of assessment (226-229). The major focus for evaluating teledermatology has been investigating whether such an approach has similar degree of diagnostic accuracy or

management concordance when compared to face to face consultations. The systematic reviews of teledermatology have all concluded that although diagnostic accuracy is lower than face to face consultations there is an acceptable degree of concordance for the management decisions for the index referral lesion (230,231). Whilst there are various proposed models of teledermatology, involving different degrees of investigation of any individual patient, in many instances clinical recommendations are made based on the examination of only the single index lesion of concern without assessment of the remainder of the patient's skin by an expert physician.

In his routine clinical practice the author has been referred patients with melanoma, that the balance of probability suggests were present at a time when those patients had been previously assessed for another cutaneous complaint, be that in primary or secondary care. Questioning of these patients revealed that many, if not most, did not receive a TBSE in either the primary or secondary care setting.

The studies presented in this chapter attempt to define the magnitude of this potential problem by identifying what proportion of melanomas are likely to be missed if physicians examined only the index lesion, rather than conducting a TBSE (total body skin examination).

10.2 Aim

Two studies were undertaken in two separate UK dermatology departments; one a district general hospital, the other a central teaching hospital on a consecutive series of patients referred to each hospitals' skin cancer clinic to identify the ratio of index to incidental melanomas. In each prospective study the numbers of patients with index melanomas (lesions for which they were referred) and any incidental melanomas identified (lesions identified during the total body skin exam) were recorded.

10.3 Materials & Methods

Subjects

Study 1:

336 patients met the inclusion criteria and agreed to participate. The mean age was 54 years (Range 16-95), and 60% were female (203/336).

Study 2:

1515 patients attended the skin lesion clinics over the 6-month period. The mean age was 57 years (Range 3-99) and 60% (912/1515) were female.

Study 1

The numbers of incidental melanomas were determined in a cohort of patients referred by primary care physicians for dermatological assessment of index lesions suspicious for any type of skin cancer (i.e. non-melanoma skin cancer and melanoma). The study was undertaken prospectively at a district general hospital that serves a defined geographic population in the West of Scotland (Stobhill Hospital, Glasgow). In this hospital the Dermatology department runs a weekly half-day skin lesion clinic, alternating between Tuesdays and Fridays, to which all referrals for the diagnosis of any type of skin cancer are assigned. All patients who attended the Friday fortnightly skin lesion clinic over a 9-month period between January and October 2010 were asked to participate as part of a larger research study involving photography of skin tumours. All participants had their index lesion recorded and photographed, along with any incidental lesions that were deemed by a consultant dermatologist to merit a biopsy. These incidental lesions were identified during a TBSE, which was routinely offered to all patients at the time of index lesion evaluation.

Study 2

In the second study, melanoma diagnoses were investigated in a hospital with a different geographical location that served a larger population. This study was undertaken prospectively at the Edinburgh Royal Infirmary, Edinburgh, a central

University teaching hospital in the South-east of Scotland. In this centre the Dermatology department runs twice weekly full-day lesion clinics, again on Tuesdays and Fridays, to which all lesion referrals from primary care are sequentially assigned. All patients who attended the Friday weekly skin lesion clinic over a 6-month period between July and December 2011 were enrolled.

In study 2 (but not Study 1) the full demographics of the total referred patient population were recorded, along with the details of all index and incidental melanomas identified during the study period. Coupled with cancer registry data produced by the Information Services Division (ISD) for the same area (South-East Scotland), it was possible to estimate how many melanomas might have been expected to occur within the referred population (4).

10.4 Results

Study 1

336 patients met the inclusion criteria and agreed to participate, which represented an enrolment rate of 87% (336/386) of the total eligible patients who attended the clinics over the study period. Patients under the age of 16 (n=7) or with a lesion on a sensitive body site (n=9) were ineligible due to ethical approval constraints applying to the photography of such lesions. The population of Study 1 were the same individuals as were sequentially photographed for the real-life image databank described in Chapter 8. Patients were excluded who failed to meet the inclusion criteria, as were the patients (n=34) who did not wish to be photographed. All of these patients had benign index lesions without any incidental malignancies. 23 additional incidental lesions were biopsied. The diagnoses and distribution of all the index lesions referred (n=336) and incidental lesions biopsied (n=23) are summarised below in Table 10.4a. The majority of the index lesions referred were not malignant 85% (287/336), although a further 15% (51/336) of these referrals were for what might best be termed “pre-malignant or dysplastic” conditions (intra-epithelial carcinomas, actinic keratoses, dysplastic nevi and keratoacanthomas).

Diagnosis	No. Index lesions n=336 (%)	No. Incidental lesions n=23 (%)
Actinic keratosis (AK)	18 (5.4)	
Angiosarcoma	1 (0.3)	
Adenexal tumour - benign	5 (1.5)	1 (4.3)
Basal cell carcinoma (BCC)	36 (10.7)	7 (30.4)
Chondrodermatitis nodular helicis (CDNH)	5 (1.5)	
Cysts - benign	4 (1.2)	
Dermatofibroma	7 (2.1)	
Haemangioma	6 (1.8)	
Intra-epithelial carcinoma	21 (6.2)	4 (17.4)
Keratoacanthoma	1 (0.3)	
Lentigo	16 (4.8)	
Lichenoid keratosis	6 (1.8)	
Melanocytic naevus- benign	85 (25.3)	
Melanocytic naevus- dysplastic	11 (3.3)	6 (26.1)
Melanoma/Melanoma in situ (MIS)	4 (1.2)	5 (21.7)
Pyogenic granuloma	5 (1.5)	
Seborrheic keratosis (Seb. K)	69 (20.5)	
Squamous cell carcinoma (SCC)	8 (2.4)	
Viral Wart	9 (2.7)	
Other (dermatoses- not lesions)	19 (5.7)	
Distribution	No. Index lesions n=336 (%)	No. Incidental lesions n=23 (%)
Lip	5 (1.5)	
Ear	12 (3.6)	
Nose	18 (5.4)	2 (8.7)
Other facial	70 (20.8)	5 (21.7)
Scalp	16 (4.8)	
Neck	9 (2.7)	
Trunk	50 (14.9)	2 (8.7)
Back	73 (21.7)	7 (30.4)
Upper limb	35 (10.4)	
Lower limb	48 (14.3)	7 (30.4)

Table 10.4a: The diagnoses and distributions of all referred index lesions (n=336) and all biopsied incidental lesions (n=23) from Study 1.

In total 9 melanomas were diagnosed in the study cohort. Four of these melanomas were index lesions, resulting in a pick-up rate of 1.2% (4/336) of the referred index

lesions. None of these index melanomas were melanoma in-situ (MIS), and their mean Breslow depth was 0.42mm- see Figure 10.4b.

Index melanomas (n=4)

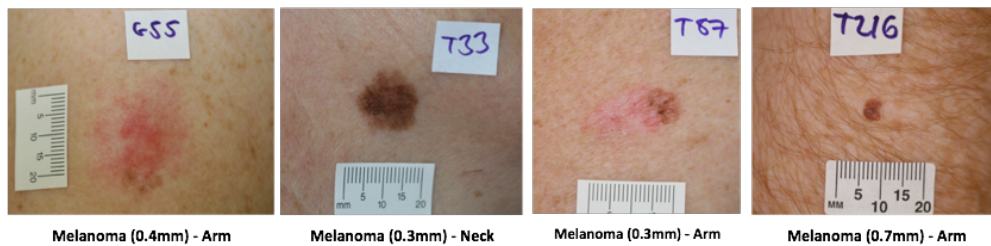


Figure 10.4b: The digital images of the four index melanomas from Study 1, with their corresponding Breslow depth in millimetres.

Five incidental melanomas were identified during TBSEs of the study cohort, equating to a pick-up rate of 1.5% (5/336). Three of these incidental melanomas were melanoma in situ (MIS), and the mean Breslow depth of the two invasive melanomas was 0.32mm - see Figure 10.4c.

Incidental melanomas (n=5)

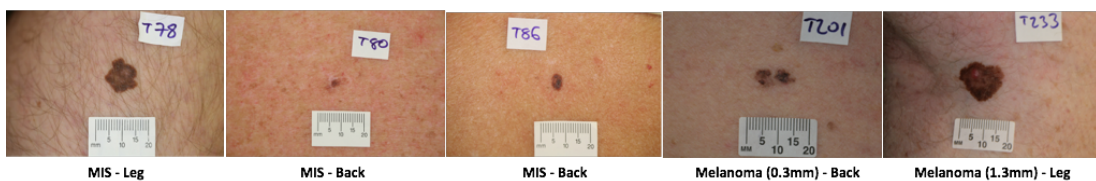


Figure 10.4c: The digital images of the five incidental melanomas from Study 1, with their corresponding Breslow depth in millimetres.

Therefore the index to incidental melanoma ratio was 4:5, with 56% of the melanomas in the study cohort detected incidentally. In all five of these incidental melanomas the index lesion had proven benign. See Figures 10.4d & Table 10.4e for

images and characteristics of the five incidental melanomas identified during study one, along with the corresponding details of their referred index lesion.

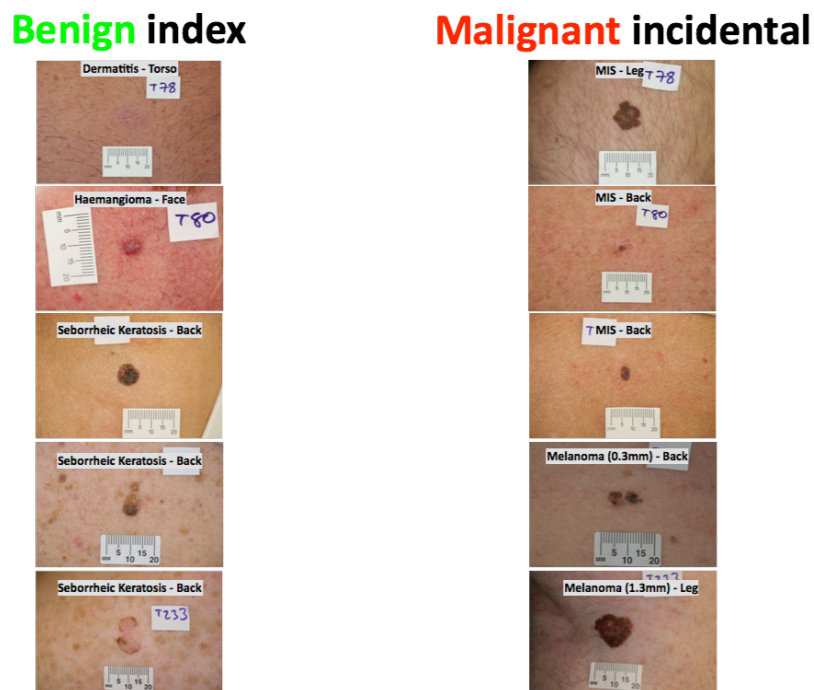


Figure 10.4d: The digital images of the five incidental melanomas and their corresponding benign index lesions from Study 1, with their corresponding Breslow depth in millimetres.

Incidental Melanomas			Referred index lesion	
No.	Breslow depth	Site	Diagnosis	Site
1	MIS	Leg	Dermatitis	Torso
2	MIS	Back	Haemangioma	Face
3	MIS	Back	Seborrheic Keratosis	Back
4	0.3mm	Back	Seborrheic Keratosis	Back
5	1.3mm	Leg	Seborrheic Keratosis	Back

Table 10.4e: The characteristics of the five incidental melanomas identified during Study 1, with their corresponding Breslow depth in millimetres. Beside these incidental melanomas are the details of the final clinical diagnosis corresponding to the referred index lesion.

Abbreviations as per Table 10.4a.

Study 2

1515 patients attended the skin lesion clinics over the 6-month period. The study population, sorted by age and sex, is graphically presented below in Figure 10.4f.

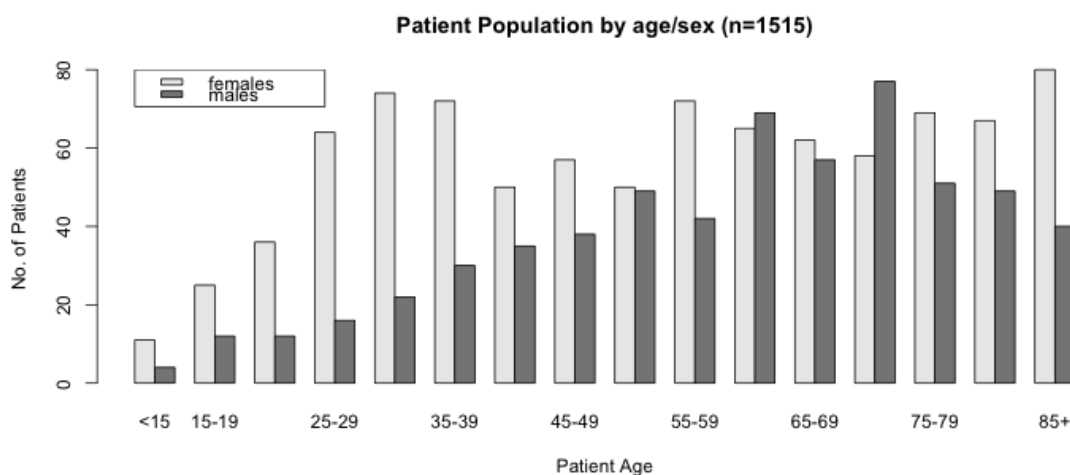


Figure 10.4f: Graph detailing the patient demographics of the Study 2 population.

In total 29 melanomas were diagnosed in the referred population. Twenty of these melanomas were index lesions, resulting in an index melanoma rate of 1.3% (20/1515) of the referred lesions. 30% (6/20) of these index melanomas were MIS, and the mean Breslow depth of the fourteen invasive melanomas was 1.03mm.

Nine incidental melanomas were identified by TBSE in the referred patients, with a pick-up rate of 0.6% (9/1515) of patients referred. Again, a third (3/9) of these incidental melanomas were MIS, and the mean Breslow depth of the six invasive melanomas was 0.47mm. The index to incidental melanoma ratio was 20:9, with 31% of melanomas diagnosed as incidental lesions. In the majority (5/9) of these incidental melanomas the index lesion was non-malignant. The exact make-up of the sites and diagnoses of the incidental melanomas and their corresponding index lesions are detailed in Table 10.4g below.

Incidental Melanomas			Referred index lesion	
No.	Breslow depth	Site	Diagnosis	Site
6	MIS	Arm	AK	Face
7	MIS	Neck	Seborrheic Keratosis	Back
8	MIS	Arm	Cyst	Back
9	0.24mm	Arm	CDNH	Ear
10	0.43mm	Back	BCC	Cheek
11	0.5mm	Arm	SCC	Scalp
12	0.66mm	Back	BCC	Back
13	0.8mm	Leg	Melanocytic naevus	Leg
14	1.6mm	Back	MIS	Cheek

Table 10.4g: *The characteristics of the nine incidental melanomas identified during Study 2, with their corresponding Breslow depth in millimetres. Beside these incidental melanomas are the details of the final clinical diagnosis corresponding to the referred index lesion.*

Abbreviations as per Table 10.4a.

Based on age and sex standardised melanoma incidence rates for South East Scotland (Information Services Division, ISD, NHS Scotland) a cohort of patients of this size and demographics participating in study two, were they members of the general population rather than persons deemed to be at high risk of skin cancer, would expect to yield an incidence figure of 0.58 invasive melanomas over a 12-month period. The number of invasive melanomas identified in this high-risk group over a 6-month period was 20, however, of which six were identified as incidental only because of TBSE.

10.5 Discussion

The data clearly show that a substantial proportion of the melanomas detected at these skin lesion clinics are incidental findings discovered because a TBSE was performed in addition to examination of the referred index lesion. Such incidental diagnoses make a significant contribution to the total number of melanoma diagnoses made. Single index lesion assessment resulted in a melanoma pick up rate of 1.3% (24/1851) compared to 2.1% (38/1851) for the assessment incorporating a total body skin examination, due to an incidental melanoma rate of 0.8% (14/1851). Failure to perform a TBSE would lead to over one third of melanomas being missed.

Comprising a total of 1851 patients consecutively recruited, these two studies are the largest published cohort(s) that prospectively investigated the rate of incidental melanomas in patients referred to skin lesion clinics. Furthermore, despite the obvious geographical and demographic differences between these two UK studies and those previously published (see Table 10.5a below), the ratios of incidental to index melanomas at 5:4 and 9:20 are in accord with those found in other of what might be considered 'high risk' populations (232-234).

Study design	Location	Population	Index : Incidental melanoma ratio	Percentage Incidental melanomas
Prospective cohort	Glasgow, UK (Study One)	n=336 skin lesion patients	4 : 5	56%
Prospective cohort	Edinburgh, UK (Study Two)	n=1515 skin lesion patients	20 : 9	31%
Retrospective cohort	Connecticut, USA (234)	n=400 skin lesion patients	5 : 6	55%
Retrospective case series	Perth, Australia (232)	n=94 melanoma patients	37 : 57	61%
Retrospective case series	Florida, USA (233)	n=126 melanoma patients	55 : 71	56%

Table 10.5a: Details of the three studies identified in the literature where comparisons were made between the rates of 'incidental' and 'index' melanomas. Also included in the table are the details of the two studies described in this chapter.

In keeping with previous UK studies most of the patients referred to the lesion clinics did not have malignant lesions (235). In addition, the majority 71% (10/14) of those with incidental melanomas were referred with lesions that were themselves benign. Relying therefore on assessment of index lesions alone, without a supplementary TBSE performed by a dermatologist (as is the norm for teledermatology) a significant proportion of the melanomas 37% (14/38) would have been returned to primary care without further examination. Had the investigative strategy sometimes suggested of only undertaking a TBSE on patients with a cancerous index lesion have applied (47), only 29% (4/14) of the incidental melanomas would have been identified. It is not clear whether the referring physicians failed to identify these incidental melanomas because they did not perform a TBSE, or because the incidental melanomas were misdiagnosed as benign, but four of the fourteen incidental melanomas identified were likely to have been within the referring physician's field of view at the time of initial index lesion examination (see Tables 10.4e & 10.4g - melanoma numbers. 3, 4, 12 & 13).

The observation that many of the incidental melanomas were located alongside benign index lesions reinforces the investigations presented elsewhere in this thesis showing that non-specialist assessment of skin lesions requires improvement, particularly as the referring physicians self-evidently did not possess the necessary skills required to consistently choose the single most relevant index lesion to submit for expert assessment.

The value of routinely including a TBSE within general dermatology consultations, even when the patient's presenting complaint does not specifically require such a complete full body examination has been reported previously, with incidental melanoma pick up rates of 0.3% and 0.6% (236,237). The crucial differences between these studies and those presented in this chapter are that the studies presented were exclusively on patients referred because of concern over the possibility of skin cancer, and because these studies specifically investigated the differences in melanoma pick up rates between single lesion assessment and TBSE. Although the apparent higher rate of incidental melanomas in the cohort(s) presented might be due to demographic differences between the patient populations, for the reasons already discussed it is

unsurprising that patients referred with suspicious index skin lesions have a higher rate of incidental melanomas than the general dermatological patient population.

A recent, frequently cited article suggesting caution in the use of TBSE is the 2016 Hoorens study comparing TBSE and lesion directed screening (LDS) (47). This study of 9325 (TBSE) and 9482 (LDS) subjects concluding that although TBSE yielded a higher absolute number of skin cancers than LDS (39 vs 8), the detection rates were similar between the two methods (2.3% vs 3.2%). On the basis that LDS was 5.6 times less time consuming (41 secs vs 232 secs) they concluded that LDS was an acceptable alternating screening method in healthcare systems with limited budgets or long waiting lists. Several concerns need to be addressed before those conclusions can be accepted. Primarily it is questionable that the time advantage claimed for LDS over TBSE is realistic given that any examination time is likely to be a small fraction of the total consultation time. Although the populations offered screening were substantial, the uptake was relatively limited with only 17.9% (1668/9325) of the TBSE cohort and 3.3% (314/9482) of the LDS cohort actually examined. The melanoma detection rate for the TBSE cohort was 0.5% (8/1668), which was lower than in the studies presented in this chapter, but these were individuals invited for screening rather than patients referred with a primary lesion of concern (albeit 6.6% of Hoorens' TBSE cohort stated that the reason for accepting the offer of screening was because they had a lesion of concern). The melanoma detection rate in the LDS cohort was 0.4% (1/248), but even in this LDS group 17.8% attended for general screening rather than truly having a lesion of concern, furthermore only 90.1% of the LDS cohort underwent a full TBSE to look for incidental lesions. It is worth noting, however, that although the numbers were small, 30% of those in the LDS group whose index lesion was deemed to be suspicious had an incidental malignancy- supporting the conclusions of the studies presented in this chapter. Further direct comparisons are not possible as the demographics of the study populations in Scotland and Belgium are likely to differ, but interestingly, in a follow up paper (in which Hoorens was a contributor) the findings were tweaked slightly with the authors concluding that due to the low participation rates they now felt that TBSE is more cost effective than LDS (46).

The natural history of melanomas is unknown, with some suggesting that lesion clinics may in effect only be screening the worried well or picking up biologically inactive malignancies (25,26,79,80,235). Nonetheless, given the importance attached to early diagnosis in melanoma, if any of the incidental melanomas detected through TBSE had progressed any delay in their diagnosis is likely to have adversely affected at least several of the patients' outcomes. Albeit not achieving formal statistical significance, in the studies presented in this chapter the Breslow thickness of the incidental melanomas was lower than those melanomas detected as index lesions (mean 0.42mm versus 0.93mm, t-test $p=0.08$). Such findings are in keeping with previous work which has demonstrated thinner Breslow depths at the time of melanoma diagnosis in those patients who have undergone a recent TBSE (238).

A further finding worthy of comment, which is germane to discussions concerning the importance of performing TBSE and screening of those who do not present to their medical practitioner with a suspicious lesion, was the additional work in Study 2, where age and sex data were used to approximate how many melanomas might have been expected in a general population of the same size and demographics to this patient study group. This calculation is not straightforward for several reasons; including relating the number of incidental lesions present at the single point in time when the patients were examined, to the rate of appearance of incidental lesions over a period during which they may present to medical services, takes no account of how long lesions are present on any individual before they choose to present to the medical services. Such considerations influence the 'pool' of subclinical disease existing (i.e. prevalence) at any single point of time in the community, which incidence rates do not capture. Furthermore, as indicated above, the patient group examined are likely to be at a higher risk for melanoma because many had been referred with other suspected skin cancer lesions, despite only a minority having been found to have skin cancer. A difference in baseline skin cancer incidence in such a selected population might be expected to be higher than the general population and such matters are not adequately reflected by age and sex standardisation techniques, but hopefully the results presented in this chapter will inform the work of others.

In summary, the finding of six incidental invasive melanomas at a particular time point against an expected total melanoma count of 0.6 over a one year period in the patients represents a tenfold difference, suggesting the presence of a significant pool of unidentified lesions being present in the community. Although the natural history of these lesions with their propensity to metastasis is unknown, the implication of these findings will be discussed more fully in Chapter 11.

Chapter 11

Summary and Conclusions

Notwithstanding recent advances in molecular targeted therapies for advanced melanoma (239,240), early detection and excision is, as yet, the only proven method of reducing mortality. The problem remains significant with approximately 2459 deaths from melanoma in the UK in 2014 (3).

Screening for both melanoma and other skin malignancies with a view to early detection is an attractive proposition, but remains controversial due to the cost implications and constraints related to the limited availability of expert assessors. As discussed in Chapter 10, if screening is to be undertaken it should be TBSE rather than LDS. The natural history of melanoma is unknown so it is not possible to be certain that malignant lesions identified and removed during screening would, if ignored, progress and become life threatening. As most skin malignancies have a latent pre-invasive phase, if public screening is to be undertaken the optimum time interval between screening must be determined. Although Pil et al. have proposed once in a lifetime screening of adults for skin malignancies (46), presumably those at greater risk might benefit from more frequent screening and screening those at low risk may not be cost effective.

Irrespective of these screening considerations, the health benefits of early removal of frankly aggressively malignant lesions are difficult to challenge (both in terms of gain in length/quality of life and avoidance of medical costs associated with advanced disease). The development of malignancy in skin lesions is usually associated with detectable change and it is likely that public education over the implications of such changes and maximising the competence of the primary assessor will be prove to be beneficial in identifying malignant lesions at the earliest opportunity.

The work presented in Chapters 3 and 4 supports the hypothesis that accurate identification of skin lesion occurs predominantly by NAPR processing and that APR, which predominates in most current public education programmes, is not of

significant utility. The acquisition of effective NAPR depends upon exposure to appropriate examples. Unfortunately, as the studies reported in Chapter 5 indicate there is inadequate exposure to appropriate dermatological lesions in the undergraduate training of medical students. In the current NHS structure, many of these students will, after graduation, become primary care practitioners and often, without further dermatological training, act as the gatekeepers to cutaneous cancer pathways. It was disappointing, therefore, that their undergraduate training failed to meet the BAD guidelines.

The work in Chapter 6 demonstrates that the diagnostic acumen of non-experts can be enhanced by repeated visual exposure to increasing numbers of images of lesions exhibiting multiple variations. The increased expertise of dermatologists over non-dermatologists identified by Chen et al. in a systematic review probably reflects their increased exposure to multiple exemplars (90). The variation in diagnostic ability amongst experts may also reflect varying exposure but even those with extensive experience can never achieve 100% diagnostic accuracy. The work in Chapter 7, notwithstanding that the learning potential of the technique was not fully evaluated, suggests that acquisition of diagnostic skills could be further improved by the introduction of 3D models.

Modern technology allows widespread distribution of high quality coloured images at low cost, but as the work presented in Chapter 8 shows, if these are restricted to “classic” images they may lead to overconfidence and the rejection of atypical presentations of malignancies as non-malignant. It is hence essential that all those responsible for initial screening of potential malignancies have exposure to as wide a variety as possible of clinical exemplars.

In the current model of care a major inhibition to the introduction to population based screening is the availability of dermatologists to undertake the expert assessments (41); rendering it essential, so far as is possible, to screen out obviously non-malignant lesions prior to specialist assessment. Irrespective of the constraints on their time it is unlikely that even with an enriched education a general medical practitioner will acquire sufficient expertise to confidently distinguish malignant from non-malignant

lesions, and as the work in Chapter 5 suggests whatever level of competency is acquired by specialist training that level of competence is likely to diminish without reinforcement.

The studies presented in Chapters 3 and 9 demonstrate that non-medically trained volunteers, by their possession of native NAPR, are capable of effectively categorising skin lesions. A practical solution resulting from this finding may be to train a cadre of non-medically qualified experts who are totally dedicated to such assessments, especially if population based screening were to be extended.

Attempts have been made to educate the public to distinguish malignant from non-malignant lesions using both NAPR and APR but apart from establishing that APR is of very little value (126), practicable methods of using NAPR by public exposure to images have yet to be devised (241,242). Although the outcomes of strategies for the education of the public are still in the process of evaluation the electronic tools described in this thesis merit further development to evaluate whether they may eventually prove to be an effective extension of the posters and leaflets that are presently used to educate the public of the warnings signs of skin cancer. Despite some justifiable anxiety as whether such approaches are safe, it is worth noting that 80% of internet users have already undertaken health related searches (243).

If, however, the use of APR is questionable and further studies suggest that the volume of images required to make public education through NAPR strategies are excessive, the most effective means of achieving public awareness will have to be determined. The likely pragmatic approach, already suggested by many dermatologists, is to advise that all changing lesions require expert evaluation (127,152,153). As common benign lesions mimicking melanoma (e.g. seborrheic keratoses) can often be pruritic particularly as they develop (and hence change) it is again questionable whether there is sufficient dermatological expertise to assess all the changing lesions that an effective public awareness campaign is likely to evoke. If the funding were made available for the training of a cadre of the specialist examiners postulated above, it might be possible for such cadre to undertake the initial assessments of all the newly changing lesions that any such public awareness campaigns would uncover. Initially limited

trials of such a system would be required to ensure cost effectiveness before it was extended widely.

Finally, whatever the insights the work presented in this thesis in respect of different diagnostic and screening strategies, the work in Chapter 9 establishes that a properly catalogued image bank is likely to be a useful diagnostic support tool for non-experts. Ultimately if a large enough bank of images could be established with appropriate software for effective cataloguing and it was possible to make this library accessible to the public it may result in earlier cancer diagnosis. To date, following on from the work described in this thesis, an iPad application has been developed and licensed to Simedics Ltd (UK) and is now available to download at <http://www.dermofit.org> (244). The Dermofit iPad app has over 1300 lesions of the most common ten diagnostic classes, enabling the user both to attempt lesion matching for diagnostic purposes and access a separate tutorial function employing multiple training images.

Several techniques have been developed to assist in the diagnosis of cutaneous lesions, the mostly widely used of which are dermoscopy and spectroscopy based (48,49). Such devices when linked to a video microscope for computer analysis have often been termed “mole scanners”. Is it possible therefore that advancing technology will render the clinical assessment of lesions by an expert unnecessary and negate the requirement for dermatological training?

Studies assessing automated computer diagnosis of skin cancers have been reported for over twenty years, and although these have shown increasing levels of experimental accuracy this has only been within a narrow spectrum of disease (245,246). The results when applied in a clinical setting have been less convincing (247-249). Perhaps the reason for their failure is that most such systems have focused on extracting and assessing individual features, so resemble a computer based APR technique. The Dermofit project differed in that it attempted to use informatics techniques to order images in the hope of replicating the expert’s internal library against which an index lesion could be referenced by a non-expert, thus mimicking an expert’s NAPR based diagnosis.

Informatics techniques have now developed and migrated from feature based algorithms to embrace machine learning (250). Recently the technology has further advanced and in February 2017 Esteva et al published a study in Nature describing the application of deep neural networks to an image bank which showed classification on a par with expert dermatologists (251). This study acknowledges and refers to the Dermofit project and includes many of the Edinburgh Dermofit library images. In simplistic terms the algorithms utilised by Esteva, which integrate experience alone (in the form of multiple exemplars) is analogous to the computer equivalent of NAPR; their approach has led to the suggestion that the dominance of the human expert in the diagnosis of skin tumours may now be challenged (252).

Abbreviations

2D	Two dimensional
3D	Three dimensional
A	Acrochordon/skin tag
ABCD	Asymmetry, Border irregularity, Colour differentiation, Diameter
AK	Actinic keratosis
APR	Analytical pattern recognition
BAD	British Association of Dermatology
BCC	Basal cell carcinoma
BN	Benign naevus
CRUK	Cancer Research UK
D1	Day 1
D6	Day 6
D10	Day 10
DF	Dermatofibroma
DGH	District general hospital
DN	Dysplastic naevus
DPR	Differential pattern recognition
H	Haemangioma
IEC	Intra-epithelial carcinoma
IQR	Inter-Quartile range
IRAS	Integrated research application system
K	Keratinocyte
LDS	Lesion directed screening
LREC	Local research ethics committee
NAPR	Non-analytical pattern recognition
NMSC	Non-melanoma skin cancer
NRES	National research ethics service
OPR	Overall pattern recognition
P	Pigmented
PG	Pyogenic granuloma
PHP	PHP: Hypertext Pre-processor
R&D	Research and development
SCC	Squamous cell carcinoma
SHA	Strategic health authority
SK	Sebhorreic keratosis
SSI	Site specific information
TBE/TBSE	Total body examination/Total body skin examination
UK	United Kingdom
URL	Uniform Resource Locator
USA	United States of America
UV	Ultraviolet

Bibliography

1. Davidoff F. Focus on performance: the 21 century revolution in medical education. *Mens Sana Monogr. Medknow Publications*; 2008 Jan;6(1):29–40.
2. Wellcome Trust. Grants awarded [Internet] 2017. Accessed 01/04/17. Available from URL <http://wellcome.ac.uk/funding/managing-grant/grants-awarded>
3. Cancer Research UK (CRUK). CancerStats [Internet]. 2014 [Accessed 01/04/17]. Available from <http://info.cancerresearchuk.org/cancerstats>
4. Information Service Divison (Scotland): Skin Cancer statistics [Internet]. 2014 [Accessed 01/04/17]. Available from <http://www.isdscotland.org/health-topics/cancer/cancer-statistics/skin>.
5. Fitzpatrick TB. The Validity and Practicality of Sun-Reactive Skin Types I Through VI. *Arch Dermatol. American Medical Association*; 1988 Jun 1;124(6):869–71.
6. Miller DL, Weinstock MA. Nonmelanoma skin cancer in the United States: incidence. *J Am Acad Dermatol. Elsevier*; 1994 May;30(5 Pt 1):774–8.
7. Staples MP, Elwood M, Burton RC, Williams JL, Marks R, Giles GG. Non-melanoma skin cancer in Australia: the 2002 national survey and trends since 1985. *Med J Aust. 2006 Jan 2;184(1):6–10.*
8. Madan V, Lear JT, Szeimies R-M. Non-melanoma skin cancer. *Lancet. 2010 Feb 20;375(9715):673–85.*
9. Schadendorf D, Fisher DE, Garbe C, Gershenwald JE, Grob J-J, Halpern A, et al. Melanoma. *Nat Rev Dis Primers. 2015 Apr 23;1:15003.*
10. Leiter U, Garbe C. Epidemiology of melanoma and nonmelanoma skin cancer--the role of sunlight. *Adv Exp Med Biol. New York, NY: Springer New York; 2008;624(Chapter 8):89–103.*
11. Shuster S. Is sun exposure a major cause of melanoma? No. *BMJ. 2008 Jul 22;337:a764.*
12. Lindqvist PG, Epstein E, Landin-Olsson M, Ingvar C, Nielsen K, Stenbeck M, et al. Avoidance of sun exposure is a risk factor for all-cause mortality: results from the Melanoma in Southern Sweden cohort. *J Intern Med. 2014 Jul;276(1):77–86.*
13. Lindqvist PG, Epstein E, Nielsen K, Landin-Olsson M, Ingvar C, Olsson H. Avoidance of sun exposure as a risk factor for major causes of death: a competing risk analysis of the Melanoma in Southern Sweden cohort. *J Intern Med. 2016 Oct;280(4):375–87.*
14. Balch CM, Gershenwald JE, Soong S-J, Thompson JF, Atkins MB, Byrd DR, et al. Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol. 2009 Dec 20;27(36):6199–206.*
15. Breslow A. Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. *Ann Surg. Lippincott, Williams, and Wilkins; 1970 Nov;172(5):902–8.*

16. Med-Art: Breslow depth with 10 year survival [Internet]. 2001 [Accessed 01/04/17]. Available from http://www.med-ars.it/galleries/various_2.htm
17. Robinson JK, Dahiya M. Basal cell carcinoma with pulmonary and lymph node metastasis causing death. *Arch Dermatol*. 2003.
18. Sahl WJ. Basal cell carcinoma: influence of tumor size on mortality and morbidity. *Int J Dermatol*. 1995 May;34(5):319–21.
19. Veness MJ. High-risk cutaneous squamous cell carcinoma of the head and neck. *J Biomed Biotechnol*. 2007;2007(3):80572.
20. Erdmann F, Lortet-Tieulent J, Schüz 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? *Int J Cancer*. 2013 Jan 15;132(2):385–400.
21. Lomas A, Leonardi-Bee J, Bath-Hextall F. A systematic review of worldwide incidence of nonmelanoma skin cancer. *Br J Dermatol*. 2012 May;166(5):1069–80.
22. American Cancer Society. Skin Cancer [Internet]. 2015 [Accessed 01/04/17]. Available from http://www.cancer.org/cancer/skin_cancer.html
23. Saladi RN, Persaud AN. The causes of skin cancer: a comprehensive review. *Drugs Today*. 2005 Jan;41(1):37–53.
24. Reinau D, Surber C, Jick SS, Meier CR. Epidemiology of basal cell carcinoma in the United Kingdom: incidence, lifestyle factors, and comorbidities. *Br J Cancer*. 2014 Jul 8;111(1):203–6.
25. Shuster S. Malignant melanoma: how error amplification by screening creates spurious disease. *Br J Dermatol*. 2009 Nov;161(5):977–9.
26. Levell NJ, Beattie CC, Shuster S, Greenberg DC. Melanoma epidemic: a midsummer night's dream? *Br J Dermatol*. 2009 Sep;161(3):630–4.
27. US Department of Health and Human Services. The Surgeon General's Call to Action to Prevent Skin Cancer. Washington (DC): Office of the Surgeon General (US); 2014.
28. Mudigonda T, Pearce DJ, Yentzer BA, Williford P, Feldman SR. The economic impact of non-melanoma skin cancer: a review. *J Natl Compr Canc Netw*. 2010 Aug;8(8):888–96.
29. Cakir BÖ, Adamson P, Cingi C. Epidemiology and economic burden of nonmelanoma skin cancer. *Facial Plast Surg Clin North Am*. 2012 Nov;20(4):419–22.
30. Guy GP, Ekwueme DU, Tangka FK, Richardson LC. Melanoma treatment costs: a systematic review of the literature, 1990-2011. *Am J Prev Med*. Elsevier; 2012 Nov;43(5):537–45.
31. Chen JT, Kempton SJ, Rao VK. The Economics of Skin Cancer. *Plastic and Reconstructive Surgery - Global Open*. 2016 Sep;4(9):e868–7.
32. Doran CM, Ling R, Byrnes J, Crane M, Searles A, Perez D, et al. Estimating the economic costs of skin cancer in New South Wales, Australia. *BMC Public*

Health. *BioMed Central*; 2015 Sep 23;15(1):952.

33. Morris S, Cox B, Bosanquet N. Cost of skin cancer in England. *Eur J Health Econ*. 2009 Jul;10(3):267–73.
34. Vallejo-Torres L, Morris S, Kinge JM, Poirier V, Verne J. Measuring current and future cost of skin cancer in England. *J Public Health (Oxf)*. 2014 Mar;36(1):140–8.
35. Rigel DS, Carucci JA. Malignant melanoma: prevention, early detection, and treatment in the 21st century. *CA Cancer J Clin*. 2000 Jul;50(4):215–36–quiz237–40.
36. Loescher LJ, Janda M, Soyer HP, Shea K, Curiel-Lewandrowski C. Advances in skin cancer early detection and diagnosis. *Semin Oncol Nurs*. 2013 Aug;29(3):170–81.
37. Linden KG. Screening and early detection of skin cancer. *Curr Oncol Rep*. 2004 Nov;6(6):491–6.
38. Detection and treatment of nonmelanoma skin cancer prevent disfigurement and death. *Cutis*. 1999 Jun;344.
39. Haddix AC, Teutsch SM, Corso PS Eds. *Prevention Effectiveness: A guide to decision analysis and economic evaluation*. New York: Oxford University Press, 2003.
40. Gordon LG, Rowell D. Health system costs of skin cancer and cost-effectiveness of skin cancer prevention and screening: a systematic review. *Eur J Cancer Prev*. 2015 Mar;24(2):141–9.
41. Robinson JK, Halpern AC. Cost-effective Melanoma Screening. *JAMA Dermatol*. 2016 Jan;152(1):19–21.
42. Boniol M, Autier P, Gandini S. Melanoma mortality following skin cancer screening in Germany. *BMJ Open*. 2015 Sep 15;5(9):e008158.
43. Katalinic A, Waldmann A, Weinstock MA, Geller AC, Eisemann N, Greinert R, et al. Does skin cancer screening save lives?: an observational study comparing trends in melanoma mortality in regions with and without screening. *Cancer*. 2012 Nov 1;118(21):5395–402.
44. Brunssen A, Waldmann A, Eisemann N, Katalinic A. Impact of skin cancer screening and secondary prevention campaigns on skin cancer incidence and mortality: A systematic review. *J Am Acad Dermatol*. Elsevier; 2017 Jan;76(1):129–139.e10.
45. Breitbart EW, Waldmann A, Nolte S, Capellaro M, Greinert R, Volkmer B, et al. Systematic skin cancer screening in Northern Germany. *J Am Acad Dermatol*. 2012 Feb;66(2):201–11.
46. Pil L, Hoorens I, Vossaert K, Kruse V, Tromme I, Speybroeck N, et al. Cost-effectiveness and Budget Effect Analysis of a Population-Based Skin Cancer Screening. *JAMA Dermatol*. 2016 Dec 14.
47. Hoorens I, Vossaert K, Pil L, Boone B, De Schepper S, Ongenaes K, et al. Total-Body Examination vs Lesion-Directed Skin Cancer Screening. *JAMA Dermatol*. 2016 Jan;152(1):27–34.

48. Mogensen M, Jemec GBE. Diagnosis of Nonmelanoma Skin Cancer/Keratinocyte Carcinoma: A Review of Diagnostic Accuracy of Nonmelanoma Skin Cancer Diagnostic Tests and Technologies. *Dermatologic Surgery*. Blackwell Publishing Inc; 2007 Oct 1;33(10):1158–74.
49. Blum A, Zalaudek I, Argenziano G. Digital Image Analysis for Diagnosis of Skin Tumors. *Seminars in Cutaneous Medicine & Surgery*. Frontline Medical Communications; 2008;27(1):11–5.
50. Jackson R. The importance of being visually literate. Observations on the art and science of making a morphological diagnosis in dermatology. *Arch Dermatol*. 1975 May;111(5):632–6.
51. Norman GR, Rosenthal D, Brooks LR, Allen SW, Muzzin LJ. The development of expertise in dermatology. *Arch Dermatol*. 1989 Aug;125(8):1063–8.
52. Brooks LR, Norman GR, Allen SW. Role of specific similarity in a medical diagnostic task. *J Exp Psychol Gen*. 1991 Sep;120(3):278–87.
53. Norman GR, Brooks LR. The Non-Analytical Basis of Clinical Reasoning. *Adv Health Sci Educ Theory Pract*. Kluwer Academic Publishers; 1997;2(2):173–84.
54. Norman G, Eva K, Brooks I, Hamstra S. Expertise in Medicine and Surgery. In: *The Cambridge Handbook of Expertise and Expert Performance* (Ericsson KA, Charness N, Hoffman RR, Feltvoich PJ, Eds). New York: Cambridge University Press, 2006; 339-354.
55. Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. *Med Educ*. 2007 Dec;41(12):1140–5.
56. Gachon J, Beaulieu P, Sei JF, Gouvernet J, Claudel JP, Lemaitre M, et al. First prospective study of the recognition process of melanoma in dermatological practice. *Arch Dermatol*. 2005 Apr;141(4):434–8.
57. Evans JSBT. Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annu Rev Psychol*. 2008 Jan;59(1):255–78.
58. Norman G. Dual processing and diagnostic errors. *Adv Health Sci Educ Theory Pract*. 2009 Sep;14 Suppl 1:37–49.
59. Friedman RJ, Rigel DS, Kopf AW. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. *CA Cancer J Clin*. 1985 May;35(3):130–51.
60. Grob JJ, Bonerandi JJ. The “ugly duckling” sign: identification of the common characteristics of nevi in an individual as a basis for melanoma screening. *Arch Dermatol*. 1998.
61. Google Images. Bond cars [Internet]. 2009 [Accessed 01/04/09]. Available from <http://www.google.co.uk>
62. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med*. 1990 Oct;65(10):611–21.
63. Schmidt HG, Rikers RMJP. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ*. 2007 Dec;41(12):1133–9.
64. McLaughlin K, Rikers RM, Schmidt HG. Is analytic information processing a

- feature of expertise in medicine? *Adv Health Sci Educ Theory Pract.* 2008 Mar;13(1):123–8.
65. Ericsson KA, Charness N, Hoffman RR, Feltvoich PJ Eds. *The Cambridge Handbook of Expertise and Expert Performance.* New York: Cambridge University Press, 2006.
 66. Patel VL, Groen GJ. Knowledge Based Solution Strategies in Medical Reasoning. *Cognitive Science.* Lawrence Erlbaum Associates, Inc; 1986 Jan 1;10(1):91–116.
 67. Pople HE. The structuring of medical diagnostics: Heuristic methods for imposing structure on ill-structured problems. In: *Artificial Intelligence in Medicine* (Szolovits P Ed). Westview Press; Colorado 1982.
 68. Rajkomar A, Dhaliwal G. Improving diagnostic reasoning to improve patient safety. *Perm J.* 2011;15(3):68–73.
 69. Lesgold A, Rubinson H, Feltovich P, Glaser R, Klopfer D, Wang Y. Expertise in a complex skill: diagnosing X-ray pictures. In: *The nature of expertise* Chi M, Glaser R, Farr M (Eds). Hillsdale; New Jersey, 1988; 311-342.
 70. Goldman L. Some investigative studies of pigmented nevi with cutaneous microscopy. *J Invest Dermatol.* 1951 Jun;16(6):407–27.
 71. Menzies SW, Zalaudek I. Why perform dermoscopy? The evidence for its role in the routine management of pigmented skin lesions. *Arch Dermatol.* 2006 Sep;142(9):1211–2.
 72. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. *The lancet oncology.* 2002.
 73. 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. *Br J Dermatol.* 2008 Sep;159(3):669–76.
 74. Binder M, Schwarz M, Winkler A. Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Archives of Dermatology* 1995.
 75. 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. *Arch Dermatol.* 2001 Oct;137(10):1343–50.
 76. Dolianitis C, Kelly J, Wolfe R, Simpson P. Comparative performance of 4 dermoscopic algorithms by nonexperts for the diagnosis of melanocytic lesions. *Arch Dermatol.* American Medical Association; 2005 Aug;141(8):1008–14.
 77. 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. *J Am Acad Dermatol.* 2007 May;56(5):759–67.
 78. Burton RC, Armstrong BK. Recent incidence trends imply a nonmetastasizing form of invasive melanoma. *Melanoma Res.* 1994 Apr;4(2):107–13.

79. Swerlick RA, Chen S. The melanoma epidemic. Is increased surveillance the solution or the problem? *Arch Dermatol*. 1996 Aug;132(8):881–4.
80. Rees JL. The melanoma epidemic: reality and artefact. *BMJ*. 1996 Jan 20;312(7024):137–8.
81. Shuster S. Diagnoses of melanoma need further investigation. *BMJ*. BMJ Group; 1996 Sep 7;313(7057):627–8.
82. Helfand M, Mahon SM, Eden KB, Frame PS, Orleans CT. Screening for skin cancer. *Am J Prev Med*. 2001 Apr;20(3 Suppl):47–58.
83. Brochez L, Verhaeghe E, Grosshans E, Haneke E, Piérard G, Ruitter D, et al. Inter-observer variation in the histopathological diagnosis of clinically suspicious pigmented skin lesions. *J Pathol*. 2002 Apr;196(4):459–66.
84. A nationwide survey of observer variation in the diagnosis of thin cutaneous malignant melanoma including the MIN terminology. CRC Melanoma Pathology Panel. *J Clin Pathol*. BMJ Group; 1997 Mar;50(3):202–5.
85. Pearl RA, Townley W, Stott D, Grobbelaar AO. Diagnosis of skin lesions by trainee surgeons: experience improves accuracy. *Ann R Coll Surg Engl*. 2009 Sep;91(6):494–9.
86. Hallock GG, Lutz DA. Prospective study of the accuracy of the surgeon's diagnosis in 2000 excised skin tumors. *Plast Reconstr Surg*. 1998 Apr;101(5):1255–61.
87. Ek EW, Giorlando F, Su SY, Dieu T. Clinical diagnosis of skin tumours: how good are we? *ANZ J Surg*. 2005 Jun;75(6):415–20.
88. Heal CF, Raasch BA, Buettner PG, Weedon D. Accuracy of clinical diagnosis of skin lesions. *Br J Dermatol*. 2008 Sep;159(3):661–8.
89. Fleischer AB, Herbert CR, Feldman SR, O'Brien F. Diagnosis of skin disease by nondermatologists. *Am J Manag Care*. 2000 Oct;6(10):1149–56.
90. Chen SC, Bravata DM, Weil E, Olkin I. A comparison of dermatologists' and primary care physicians' accuracy in diagnosing melanoma: a systematic review. *Arch Dermatol*. 2001 Dec;137(12):1627–34.
91. Osborne JE, Chave TA, Hutchinson PE. Comparison of diagnostic accuracy for cutaneous malignant melanoma between general dermatology, plastic surgery and pigmented lesion clinics. *Br J Dermatol*. 2003 Feb;148(2):252–8.
92. 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. *Br J Dermatol*. 2004 Apr;150(4):687–92.
93. 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. *Arch Dermatol*. American Medical Association; 2010 Mar 1;146(3):343–4.
94. Green AR, Elgart GW, Ma F, Federman DG, Kirsner RS. Documenting dermatology practice: ratio of cutaneous tumors biopsied that are malignant. *Dermatol Surg*. 2004 Sep;30(9):1208–9.

95. Duque MI, Jordan JR, Fleischer AB, Williford PM, Feldman SR, Teuschler H, et al. Frequency of seborrheic keratosis biopsies in the United States: a benchmark of skin lesion care quality and cost effectiveness. *Dermatol Surg.* 2003 Aug;29(8):796–801–discussion801.
96. Dinnes J, Matin RN, Moreau JF, Patel L. Tests to assist in the diagnosis of cutaneous melanoma in adults: a generic protocol. *The Cochrane Collaboration* 2015.
97. Dinnes J, Wong KY, Gulati A, Chuchu N. Tests to assist in the diagnosis of keratinocyte skin cancers in adults: a generic protocol. *The Cochrane Collaboration* 2015.
98. Gerbert B, Maurer T, Berger T, Pantilat S, McPhee SJ, Wolff M, et al. Primary care physicians as gatekeepers in managed care. Primary care physicians“ and dermatologists“ skills at secondary prevention of skin cancer. *Arch Dermatol.* 1996 Sep;132(9):1030–8.
99. Feldman SR, Fleischer AB, Chen JG. The gatekeeper model is inefficient for the delivery of dermatologic services. *J Am Acad Dermatol.* 1999 Mar;40(3):426–32.
100. Dimensional Imaging Website. URL <http://www.di3d.com/index.php>
101. Coates H. Ensuring research integrity The role of data management in current crises. *Coll res libr news.* Association of College & Research Libraries; 2014 Dec 1;75(11):598–601.
102. Edinburgh Dermofit Library. URL <http://licensing.eri.ed.ac.uk/i/software/dermofit-image-library.html>.
103. Burge S, British Association of University Teachers of Dermatology. Teaching dermatology to medical students: a survey of current practice in the U.K. *Br J Dermatol.* 2002 Feb;146(2):295–303.
104. Yaakub A, Cohen SN, Singh M, Goulding JMR. Dermatological content of U.K. undergraduate curricula: where are we now? *Br J Dermatol.* 2017 Mar;176(3):836.
105. Holme SA, Scott-Lang VE, Ooi ET, Matthews AG, Darling MI, Needham D, et al. The south-east Scotland dermatology workload study: 30 years' analysis. *Br J Dermatol.* Blackwell Publishing Ltd; 2012 Jul;167(1):123–30.
106. Interwrite Cricket. PRS-Cricket radio-based handset. URL <http://einstruction.co.uk/prs/hardware/cricket>
107. R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL <http://www.R-project.org>.
108. Torgerson WS. Multidimensional scaling: I. Theory and method. *Psychometrika.* Springer-Verlag; 1952;17(4):401–19.
109. Borg I, Groenen PJF. *Modern Multidimensional Scaling: Theory and Applications* Springer: New York, 2005.
110. Scotch Malt Whisky. URL <http://www.scotchmaltwhisky.co.uk/whiskyflavourmap.htm>

111. Likert R. A technique for the measurement of attitudes. *Archives of psychology*. 1932;22(140):5–55.
112. Cornell JA, WJ. Youden—The man and his methodology. *ASQC Statistics Division Newsletter*. 1992.
113. Ng AY, Jordan MI, Weiss Y. On spectral clustering: Analysis and an algorithm. *NIPS*. 2001.
114. Pelaccia T, Tardif J, Tribby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online*. 2011 Mar 14;16.
115. Monajemi A, Yaghmaei M. Puzzle test: A tool for non-analytical clinical reasoning assessment. *Med J Islam Repub Iran*. 2016;30:438.
116. Rigel DS, Russak J, Friedman R. The evolution of melanoma diagnosis: 25 years beyond the ABCDs. *CA Cancer J Clin*. 2010 Sep;60(5):301–16.
117. American Academy of Dermatology: ABCDs of Melanoma Detection [Internet] 2010 [Accessed 15/11/10]. Available from: <http://www.aad.org/public/exams/abcde.html>.
118. British Association of Dermatologists: Sun Awareness - mole checking [Internet]. 2010 [Accessed 15/11/10]. Available from <http://www.bad.org.uk/site/719/default.aspx>.
119. Australian College of Dermatology: A-Z of Skin- Moles & Melanoma [Internet] 2010 [Accessed 15/11/10]. Available from http://www.dermcoll.asn.au/public/a-z_of_skin-mole_melanoma.asp.
120. European Academy of Dermatology and Venereology: Moles and Malignant Melanoma Patient Information Leaflet. [Internet] 2010. [Accessed 15/11/10]. Available from <http://www.eadv.org/patient-corner/leaflets/eadv-leaflets/moles-and-malignant-melanoma-patient-information-leaflet>
121. Koh HK, Miller DR, Geller AC, Clapp RW, Mercer MB, Lew RA. Who discovers melanoma? Patterns from a population-based survey. *J Am Acad Dermatol*. 1992 Jun;26(6):914–9.
122. Temoshok L, DiClemente RJ, Sweet DM, Blois MS, Sagebiel RW. Factors related to patient delay in seeking medical attention for cutaneous malignant melanoma. *Cancer*. 1984 Dec 15;54(12):3048–53.
123. Richard MA, Grob JJ, Avril MF, Delaunay M, Gouvernet J, Wolkenstein P, et al. Delays in diagnosis and melanoma prognosis (I): the role of patients. *Int J Cancer*. 2000 May 20;89(3):271–9.
124. Richard MA, Grob JJ, Avril MF, Delaunay M, Gouvernet J, Wolkenstein P, et al. Delays in diagnosis and melanoma prognosis (II): the role of doctors. *Int J Cancer*. 2000 May 20;89(3):280–5.
125. Criscione VD, Weinstock MA. Melanoma thickness trends in the United States, 1988-2006. *J Invest Dermatol*. 2010 Mar;130(3):793–7.
126. Girardi S, Gaudy C, Gouvernet J, Teston J, Richard MA, Grob J-J. Superiority of a cognitive education with photographs over ABCD criteria in the education of the general population to the early detection of melanoma: a randomized study. *Int J*

- Cancer. 2006 May 1;118(9):2276–80.
127. Liu W, Hill D, Gibbs AF, Tempany M, Howe C, Borland R, et al. What features do patients notice that help to distinguish between benign pigmented lesions and melanomas?: the ABCD(E) rule versus the seven-point checklist. *Melanoma Res.* 2005 Dec;15(6):549–54.
 128. McGovern TW, Litaker MS. Clinical predictors of malignant pigmented lesions. A comparison of the Glasgow seven-point checklist and the American Cancer Society's ABCDs of pigmented lesions. *J Dermatol Surg Oncol.* 1992 Jan;18(1):22–6.
 129. Healsmith MF, Bourke JF, Osborne JE, Graham-Brown RA. An evaluation of the revised seven-point checklist for the early diagnosis of cutaneous malignant melanoma. *Br J Dermatol.* 1994 Jan;130(1):48–50.
 130. Thomas L, Tranchand P, Berard F, Secchi T, Colin C, Moulin G. Semiological value of ABCDE criteria in the diagnosis of cutaneous pigmented tumors. *Dermatology (Basel).* 1998;197(1):11–7.
 131. Barnhill RL, Roush GC, Ernstoff MS, Kirkwood JM. Interclinician agreement on the recognition of selected gross morphologic features of pigmented lesions. *Studies of melanocytic nevi V. J Am Acad Dermatol.* 1992 Feb;26(2 Pt 1):185–90.
 132. Gunasti S, Mulayim MK, Fettahoglu B, Yucel A, Burgut R, Sertdemir Y, et al. Interrater agreement in rating of pigmented skin lesions for border irregularity. *Melanoma Res.* 2008 Aug;18(4):284–8.
 133. Bränström R, Hedblad M-A, Krakau I, Ullén H. Laypersons' perceptual discrimination of pigmented skin lesions. *J Am Acad Dermatol.* 2002 May;46(5):667–73.
 134. Robinson JK, Turrisi R. Skills training to learn discrimination of ABCDE criteria by those at risk of developing melanoma. *Arch Dermatol.* 2006 Apr;142(4):447–52.
 135. Ross BH. Reminders and their effects in learning a cognitive skill. *Cogn Psychol.* 1984 Jul;16(3):371–416.
 136. Allen SW, Brooks LR, Norman GR, Rosenthal D. Effect of prior examples on rule-based diagnostic performance. *Res Med Educ.* 1988;27:9–14.
 137. Allen SW, Norman GR, Brooks LR. Experimental studies of learning dermatologic diagnosis: The impact of examples. 1992 Jan;4(1):35–44.
 138. Regehr G, Cline J, Norman GR, Brooks L. Effect of processing strategy on diagnostic skill in dermatology. *Acad Med.* 1994 Oct;69(10 Suppl):S34–6.
 139. Kulatunga-Moruzi C, Brooks LR, Norman GR. Using comprehensive feature lists to bias medical diagnosis. *J Exp Psychol Learn Mem Cogn.* 2004 May;30(3):563–72.
 140. Norman G. Building on experience--the development of clinical reasoning. *N Engl J Med.* 2006 Nov 23;355(21):2251–2.
 141. Norman GR, Brooks LR, Colle CL. The benefit of diagnostic hypotheses in clinical reasoning: experimental study of an instructional intervention for forward and backward reasoning. *Cognition and* 1999.

142. SkinCancerNet ("A comprehensive online skin cancer information resource developed by the AAD"). Melanoma-What it Looks Like [Internet] 2010. [Accessed 15/11/10]. Available from <http://www.skincarephysicians.com/skincancernet/melanoma.html>
143. MacKenzie-Wood AR, Milton GW, de Launey JW. Melanoma: accuracy of clinical diagnosis. *Australas J Dermatol*. 1998 Feb;39(1):31–3.
144. Marks R, Jolley D, McCormack C, Dorevitch AP. Who removes pigmented skin lesions? *J Am Acad Dermatol*. 1997 May;36(5 Pt 1):721–6.
145. Rigel DS, Friedman RJ. The rationale of the ABCDs of early melanoma. *J Am Acad Dermatol*. Elsevier; 1993 Dec 1;29(6):1060–1.
146. Abbasi NR, Shaw HM, Rigel DS, Friedman RJ, McCarthy WH, Osman I, et al. Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria. *JAMA*. 2004 Dec 8;292(22):2771–6.
147. Rigel DS, Friedman RJ, Kopf AW, Polsky D. ABCDE--an evolving concept in the early detection of melanoma. *Arch Dermatol*. American Medical Association; 2005 Aug;141(8):1032–4.
148. Zaharna M, Brodell RT. It's time for a 'change' in our approach to early detection of malignant melanoma. *Clin Dermatol*. 2003 Sep;21(5):456–8.
149. Moynihan GD. The 3 Cs of melanoma: time for a change? *J Am Acad Dermatol*. 1994 Mar;30(3):510–1.
150. Marghoob AA, Slade J, Kopf AW, Rigel DS, Friedman RJ, Perelman RO. The ABCDs of melanoma: why change? *J Am Acad Dermatol*. 1995 Apr;32(4):682–4.
151. Fox GN. ABCD-EFG for diagnosis of melanoma. *Clin Exp Dermatol*. 2005 Nov;30(6):707.
152. Weinstock MA. ABCD, ABCDE, and ABCCDEEEEFNU. *Arch Dermatol*. 2006 Apr;142(4):528.
153. Weinstock MA. Cutaneous melanoma: public health approach to early detection. *Dermatol Ther*. 2006 Jan;19(1):26–31.
154. Luttrell MJ, Hofmann-Wellenhof R, Fink-Puches R, Soyer HP. The AC Rule for melanoma: a simpler tool for the wider community. *J Am Acad Dermatol*. 2011 Dec;65(6):1233–4.
155. American Academy of Dermatology Ad Hoc Task Force for the ABCDEs of Melanoma, Tsao H, Olazagasti JM, Cordero KM, Brewer JD, Taylor SC, et al. Early detection of melanoma: reviewing the ABCDEs. *J Am Acad Dermatol*. 2015 Apr;72(4):717–23.
156. Schofield JK, Fleming D, Grindlay D, Williams H. Skin conditions are the commonest new reason people present to general practitioners in England and Wales. *Br J Dermatol*. 2011 Nov;165(5):1044–50.
157. British Association of Dermatologists: Dermatology in the Undergraduate Medical Curriculum [Internet] 2006. [Accessed 03/08/11]. Available from <http://www.bad.org.uk/site/617/default.aspx>.
158. Clayton R, Perera R, Burge S. Defining the dermatological content of the

- undergraduate medical curriculum: a modified Delphi study. *Br J Dermatol.* 2006 Jul;155(1):137–44.
159. Davies E, Burge S. Audit of dermatological content of U.K. undergraduate curricula. *Br J Dermatol.* 2009 May;160(5):999–1005.
 160. Norman GR, Streiner DL. *Biostatistics: The Bare Essentials.* Shelton: People's Medical Publishing House, 2008: p60.
 161. DermNet NZ: Facts about skin from the New Zealand Dermatological Society. URL <http://www.dermnet.org.nz>
 162. EEMeC: University of Edinburgh Electronic Medical Curriculum. Not available outside University network.
 163. Wikipedia. URL <http://www.wikipedia.org>
 164. British Association of Dermatologists Website: Undergraduate Education Section. URL <http://www.bad.org.uk/site/617/default.aspx>.
 165. Emedicine. Medscape Reference. URL <http://emedicine.medscape.com/>
 166. Primary Care Dermatology Society (PCDS). Image Atlas. URL <http://www.pcds.org.uk/pcdshome/image-atlas>.
 167. Patient.co.uk. "Comprehensive health information as provided by GPs and nurses to patients during consultations". URL <http://www.patient.co.uk/display/16777222/>
 168. Davidson's Principles and Practice of Medicine. Colledge NR, Walker BR, Ralston SH. Edinburgh: Churchill Livingstone, 2010
 169. Clinical Dermatology. Weller RPJB, Hunter JAA, Savin JA, Dahl MV. Oxford: Blackwell Publishing, 2008.
 170. Dermatology: an illustrated colour text. Gawkrödger DJ. Edinburgh: Churchill Livingstone, 2003.
 171. Oxford Handbook of Clinical Specialties. Collier J, Longmore M, Turmezei T, Mafi AR. Oxford: Oxford University Press, 2009.
 172. ABC of Dermatology. Buxton PK, Morris-Jones R. Oxford: Wiley-Blackwell, 2009.
 173. Crash course: Ophthalmology, Dermatology, ENT. Cheung ST, Islam N, Keegan D, McGuinness J, Stouthis N. St Louis: Mosby, 2009.
 174. Physical Signs in Dermatology. Lawrence CM, Cox NH. St Louis: Mosby, 2001.
 175. Fitzpatrick's Dermatology in General Medicine. Wolff K, Goldsmith LA, Katz SI, Gilcrest BA, Paller AS, Leffell DJ. Maidenhead: McGraw-Hill, 2007.
 176. Differential Diagnosis in Dermatology. Ashton R, Leppard B. Oxford: Radcliffe Publishing Ltd, 2004.
 177. Rook's Textbook of Dermatology. Burns T, Breathnach S, Cox N, Griffiths C. Oxford: Wiley-Blackwell, 2010.

178. Lecture Notes: Dermatology. Graham-Brown R, Burns T. Oxford: Wiley-Blackwell, 2010.
179. Eva KW, Cunnington JPW, Reiter HI, Keane DR, Norman GR. How can I know what I don't know? Poor self assessment in a well-defined domain. *Adv Health Sci Educ Theory Pract*. 2004;9(3):211–24.
180. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006 Sep 6;296(9):1094–102.
181. Burge SM. Learning dermatology. *Clin Exp Dermatol*. 5 ed. 2004 May;29(3):337–40.
182. Finlay AY, Coles EC, Dawber RP, Graham-Brown RA, Hunter JA, Marks JM. Dermatology examination performance: wide variation between different teaching centres. *Med Educ*. 1994 Jul;28(4):301–6.
183. Chiang YZ, Tan KT, Chiang YN, Burge SM, Griffiths CEM, Verbov JL. Evaluation of educational methods in dermatology and confidence levels: a national survey of UK medical students. *Int J Dermatol*. 2011 Feb;50(2):198–202.
184. Cliff S, Bedlow AJ, Melia J, Moss S, Harland CC. Impact of skin cancer education on medical students' diagnostic skills. *Clin Exp Dermatol*. 2003 Mar;28(2):214–7.
185. Marks J. Teaching dermatology to undergraduates in the 1980s. *Clin Exp Dermatol*. 1984 May;9(3):263–6.
186. Burge SM, Lancaster T. Assessment in undergraduate dermatology. *Clin Exp Dermatol*. 2004 Jul;29(4):441–6.
187. Hansra NK, O'Sullivan P, Chen CL, Berger TG. Medical school dermatology curriculum: are we adequately preparing primary care physicians? *J Am Acad Dermatol*. Elsevier; 2009 Jul;61(1):23–29.e1.
188. Kerr OA, Walker J, Boohan M. General practitioners' opinions regarding the need for training in dermatology at undergraduate and postgraduate levels. *Clin Exp Dermatol*. 2006 Jan;31(1):132–3.
189. Moore MM, Geller AC, Zhang Z, Hayes BB, Bergstrom K, Graves JE, et al. Skin cancer examination teaching in US medical education. *Arch Dermatol*. 2006 Apr;142(4):439–44.
190. Enk CD, Gilead L, Smolovich I, Cohen R. Diagnostic performance and retention of acquired skills after dermatology elective. *Int J Dermatol*. 2003 Oct;42(10):812–5.
191. Sherertz EF. Learning dermatology on a dermatology elective. *Int J Dermatol*. 1990 Jun;29(5):345–8.
192. Whitaker-Worth DL, Susser WS, Grant-Kels JM. Clinical dermatologic education and the diagnostic acumen of medical students and primary care residents. *Int J Dermatol*. 1998 Nov;37(11):855–9.
193. Rourke L, Oberholtzer S, Chatterley T, Brassard A. Learning to detect, categorize, and identify skin lesions: a meta-analysis. *JAMA Dermatol*. 2015 Mar;151(3):293–301.

194. Nault AM, Bhatia AC, Xu S. Accuracy and sources of images from direct Google image searches for common dermatology terms. *Cutis*. 2016 Nov;98(5):E6–E8.
195. VisualDx. Visual Clinical Decision Support System. Melanoma image [Internet] 2010. [Accessed 15/11/10]. Available from [http:// www.visualdx.com](http://www.visualdx.com).
196. Dermnet. Skin Disease Atlas. Melanoma image [Internet] 2010. [Accessed 15/11/10]. Available from <http://www.dermnet.com/images/malignant-melanoma/picture/20594>
197. Norman G, Bordage G, Page G, Keane D. How specific is case specificity? *Med Educ*. 2006 Jul;40(7):618–23.
198. Schnalke T. A brief history of the dermatologic moulage in Europe. Part I. The origin. Vol. 27, *International journal of dermatology*. 1988. 6 p.
199. Schnalke T. A brief history of the dermatologic moulage in Europe. Part II. Breakthrough and rise. *Int J Dermatol*. 1992 Feb;31(2):134–41.
200. Schnalke T. A brief history of the dermatologic moulage in Europe. Part III: Prosperity and decline. Vol. 32, *International journal of dermatology*. 1993. 11 p.
201. Garg A, Haley H-L, Hatem D. Modern moulage: evaluating the use of 3-dimensional prosthetic mimics in a dermatology teaching program for second-year medical students. *Arch Dermatol*. 2010 Feb;146(2):143–6.
202. Robinson JK, Lio P, Hernandez C, Kim NN, Lee KC, Wickless H, et al. Medical student detection of melanoma: clinical skills. *Arch Dermatol*. 2010 Oct;146(10):1175–7.
203. Hernandez C, Mermelstein R, Robinson JK, Yudkowsky R. Assessing students' ability to detect melanomas using standardized patients and moulage. *J Am Acad Dermatol*. 2013 Mar;68(3):e83–8.
204. Krishna SK. Modern moulage. *Indian J Dermatol Venereol Leprol*. 2011 Jan;77(1):64.
205. Parish LC, Geiges ML. The wax model. *Skinmed*. 2011 Sep;9(5):275–7.
206. Goulart JM, Dusza S, Pillsbury A, Soriano RP, Halpern AC, Marghoob AA. Recognition of melanoma: a dermatologic clinical competency in medical student education. *J Am Acad Dermatol*. 2012 Oct;67(4):606–11.
207. Norman G. Anatomical mysteries. *Adv Health Sci Educ Theory Pract*. 2010 May;15(2):149–51.
208. Fitzpatrick's colour atlas of clinical dermatology. Wolff K, Johnson RA. Maidenhead: McGraw-Hill, 2009.
209. *Dermatology*. Bologna JL. London: Elsevier, 2007
210. *Atlas of clinical dermatology*. du Vivier A. London: Elsevier, 2002
211. *Braun Falco's Dermatology*. Burgdorf W, Plewig G, Wolff HH, Landthaler M. New York: Springer, 2008.
212. Lloyd FJ, Reyna VF. Clinical gist and medical education: connecting the dots. *JAMA*. 2009 Sep 23;302(12):1332–3.

213. Keyzers C, Xiao DK, Földiák P, Perrett DI. The speed of sight. *Journal of cognitive neuroscience* 2001.
214. Grill-Spector K, Kanwisher N. Visual recognition: as soon as you know it is there, you know what it is. *Psychol Sci.* 2005 Feb;16(2):152–60.
215. Federman DG, Kirsner RS. The abilities of primary care physicians in dermatology: implications for quality of care. *Am J Manag Care.* 1997 Oct;3(10):1487–92.
216. Marghoob A, Kopf AW, Bart RS, Sanfilippo L, Silverman MK, Lee P, et al. Risk of another basal cell carcinoma developing after treatment of a basal cell carcinoma. *J Am Acad Dermatol.* Elsevier; 1993 Jan 1;28(1):22–8.
217. Schinstine M, Goldman GD. Risk of synchronous and metachronous second nonmelanoma skin cancer when referred for Mohs micrographic surgery. *J Am Acad Dermatol.* 2001 Mar;44(3):497–9.
218. Bergstresser PR, Halprin KM. Multiple sequential skin cancers. The risk of skin cancer in patients with previous skin cancer. *Arch Dermatol.* 1975 Aug;111(8):995–6.
219. Raasch BA, Buettner PG. Multiple nonmelanoma skin cancer in an exposed Australian population. *Int J Dermatol.* 2002 Oct;41(10):652–8.
220. Schreiber MM, Moon TE, Fox SH, Davidson J. The risk of developing subsequent nonmelanoma skin cancers. *Journal of the American Academy of Dermatology* 1990.
221. Scope A, Dusza SW, Halpern AC, Rabinovitz H, Braun RP, Zalaudek I, et al. The “Ugly Duckling” Sign: Agreement Between Observers. *Arch Dermatol.* American Medical Association; 2008 Jan 1;144(1):58–64.
222. Gaudy-Marqueste C, Wazaefi Y, Bruneu Y, Triller R, Thomas L, Pellacani G, et al. Ugly Duckling Sign as a Major Factor of Efficiency in Melanoma Detection. *JAMA Dermatol.* 2017 Apr 1;153(4):279–84.
223. Perednia DA, Brown NA. Teledermatology: one application of telemedicine. *Bull Med Libr Assoc.* 1995 Jan;83(1):42–7.
224. Burgiss SG, Julius CE, Watson HW, Haynes BK, Buonocore E, Smith GT. Telemedicine for Dermatology Care in Rural Patients. *Telemedicine Journal.* 1997 Jan;3(3):227–33.
225. Eedy DJ, Wootton R. Teledermatology: a review. *Br J Dermatol.* 2001 Apr;144(4):696–707.
226. Mahendran R, Goodfield MJD, Sheehan-Dare RA. An evaluation of the role of a store-and-forward teledermatology system in skin cancer diagnosis and management. *Clin Exp Dermatol.* 2005 May;30(3):209–14.
227. Moreno-Ramirez D, Ferrandiz L, Galdeano R, Camacho FM. Teledermatoscopy as a triage system for pigmented lesions: a pilot study. *Clin Exp Dermatol.* 2006 Jan;31(1):13–8.
228. Moreno-Ramirez D, Ferrandiz L, Nieto-Garcia A, Carrasco R, Moreno-Alvarez P, Galdeano R, et al. Store-and-forward teledermatology in skin cancer triage: experience and evaluation of 2009 teleconsultations. *Arch Dermatol.* 2007

Apr;143(4):479–84.

229. Wurm EMT, Hofmann-Wellenhof R, Wurm R, Soyer HP. Telemedicine and teledermatology: Past, present and future. *J Dtsch Dermatol Ges.* 2008 Feb;6(2):106–12.
230. Warshaw EM, Hillman YJ, Greer NL, Hagel EM, MacDonald R, Rutks IR, et al. Teledermatology for diagnosis and management of skin conditions: a systematic review. *J Am Acad Dermatol.* 2011 Apr;64(4):759–72.
231. Finnane A, Dallest K, Janda M, Soyer HP. Teledermatology for the Diagnosis and Management of Skin Cancer: A Systematic Review. *JAMA Dermatol.* 2017 Mar 1;153(3):319–27.
232. Cherian P, Tait CP. Melanoma in private practice: do dermatologists make a difference? *Australas J Dermatol.* 2009 Nov;50(4):257–60.
233. Kantor J, Kantor DE. Routine dermatologist-performed full-body skin examination and early melanoma detection. *Arch Dermatol.* 2009 Aug;145(8):873–6.
234. Viola KV, Tolpinrud WL, Gross CP, Kirsner RS, Imaeda S, Federman DG. Outcomes of referral to dermatology for suspicious lesions: implications for teledermatology. *Arch Dermatol.* 2011 May;147(5):556–60.
235. Weatherhead SC, Lawrence CM. Melanoma screening clinics: are we detecting more melanomas or reassuring the worried well? *Br J Dermatol.* 2006 Mar;154(3):539–41.
236. 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. *J Am Acad Dermatol.* 2012 Feb;66(2):212–9.
237. Moran B, McDonald I, Wall D, O’Shea SJ, Ryan C, Ryan AJ, et al. Complete skin examination is essential in the assessment of dermatology patients: findings from 483 patients. *Br J Dermatol.* 2011 Nov;165(5):1124–6.
238. Aitken JF, Elwood M, Baade PD, Youl P, English D. Clinical whole-body skin examination reduces the incidence of thick melanomas. *Int J Cancer.* 2010 Jan 15;126(2):450–8.
239. van Zeijl MCT, van den Eertwegh AJ, Haanen JB, Wouters MWJM. (Neo)adjuvant systemic therapy for melanoma. *European Journal of Surgical Oncology.* Elsevier; 2017 Mar 1;43(3):534–43.
240. D Aniello C, Perri F, Scarpati G, Pepa C, Pisconti S, Montesarchio V, et al. Melanoma Adjuvant Treatment: current insight and clinical features. *CCDT.* 2017;17(999):1–1.
241. Robertson K, McIntosh RD, Bradley-Scott C, MacFarlane S, Rees JL. Image training, using random images of melanoma, performs as well as the ABC(D) criteria in enabling novices to distinguish between melanoma and mimics of melanoma. *Acta Derm Venereol.* 2014 May;94(3):265–70.
242. Cornell E, Robertson K, McIntosh RD, Rees JL. Viewing Exemplars of Melanomas and Benign Mimics of Melanoma Modestly Improves Diagnostic Skills in Comparison with the ABCD Method and Other Image-based Methods for Lay Identification of Melanoma. *Acta Derm Venereol.* 2015 Jul;95(6):681–5.

243. White RW, Horvitz E. Cyberchondria. *ACM Trans Inf Syst.* 2009 Nov 1;27(4):1–37.
244. Simedics Ltd. Dermofit: Professional training and digital reference tool to aid skin lesion identification and diagnosis. URL <http://www.dermofit.org>
245. Rosado B, Menzies S, Harbauer A, Pehamberger H, Wolff K, Binder M, et al. Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis. *Arch Dermatol.* 2003 Mar;139(3):361–7–discussion366.
246. Masood A, Al-Jumaily AA. Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. *Int J Biomed Imaging.* 2013;2013:323268.
247. Boldrick JC, Layton CJ, Nguyen J, Swetter SM. Evaluation of digital dermoscopy in a pigmented lesion clinic: clinician versus computer assessment of malignancy risk. *J Am Acad Dermatol.* 2007 Mar;56(3):417–21.
248. Vestergaard ME, Menzies SW. Automated Diagnostic Instruments for Cutaneous Melanoma. *Seminars in Cutaneous Medicine and Surgery. Frontline Medical Communications;* 2008;27(1):32–6.
249. 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 Res.* 2009 Jun;19(3):180–4.
250. Li L, Zhang Q, Ding Y, Jiang H, Thiers BH, Wang JZ. Automatic diagnosis of melanoma using machine learning methods on a spectroscopic system. *BMC Med Imaging.* 2014 Oct 13;14:36.
251. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017 Feb 2;542(7639):115–8.
252. Rees JL. Software is Eating the Clinic. *Acta Derm Venereol.* 2017 Apr 6;97(4):423–4.

List of the publications, abstracts and presentations arising from the author's work on the Dermofit project

Publications

1. 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 Derm Venereol.* 2013; 93(6): 689-92. PMID 23695107.
2. Ballerini L, Fisher RB, Aldridge B, Rees J, 2012. A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. In: ME Celebi, G Schaefer (Eds.) *Color Medical Image Analysis*, Springer (London) pp. 63-86. ISBN 9789400753884.
3. Aldridge RB, Maxwell SS, Rees JL. Dermatology undergraduate skin cancer training: A disconnect between recommendations, clinical exposure and competence. *BMC Medical Education* 2012; 12:27. PMID: 22569037.
4. Ballerini L, Fisher RB, Aldridge B, Rees J., 2012. Non-Melanoma skin lesion classification using colour image data in a hierarchical K-NN classifier. In: A Frangi, A Santos (Eds.) *Proc. International Symposium on Biomedical Imaging (ISBI): From Nano to Macro*, IEEE pp. 358 to 361. ISBN: 978145771857.
5. Aldridge RB, Glodzik D, Ballerini L, Fisher RB, Rees JL. Utility of non-rule-based visual matching as a strategy to allow novices to achieve skin lesion diagnosis. *Acta Derm Venereol.* 2011; 91(3): 279-83. PMID: 21461552.
6. Aldridge RB, Zanotto M, Ballerini L, Fisher RB, Rees JL. Novice identification of melanoma: not quite as straightforward as the ABCDs. *Acta Derm Venereol.* 2011; 91(2): 125-30. PMID: 21311845.
7. Li X, Aldridge B, Rees J, Fisher R., 2011. Estimating the ground truth from multiple individual segmentations incorporating prior pattern analysis with application to with application to skin lesion segmentation. In: S Wright, X Pan, M Liebling (Eds.) *Proc. International Symposium on Biomedical Imaging (ISBI): From Nano to Macro*, IEEE pp. 1438 to 1441. ISBN: 9781424441273.
8. Zanotto M, Ballerini L, Fisher RB, Aldridge B, Rees J., 2011. Visual Cues Do Not Improve Lesion ABC(D) Grading. In: DJ Manning, CK Abbey (Eds.) *Proc. Medical Imaging 2011: Image Perception, Observer Performance, and Technology Assessment*, SPIE vol. 7966. Bellingham: SPIE. pp. 79660U-1 to 79660U-10. ISBN: 9780819485083.
9. Li X, Aldridge B, Rees J, Fisher R., 2010. Estimating the ground truth from multiple individual segmentations with application to skin lesion segmentation. In: AH Bhalerao, NM Rajpoot (Eds.) *Proc. Medical Image Understanding and Analysis (MIUA)*, BMVA Press. pp. 101 to 106. ISBN: 9780956615008.
10. Aldridge RB, Li X, Ballerini L, Fisher RB, Rees JL. Teaching dermatology using 3-dimensional virtual reality. *Arch Dermatol.* 2010; 146(10):1184-5; PMID: 20956667.
11. Laskaris N, Ballerini L, Fisher RB, Aldridge B, Rees J., 2010. Fuzzy description of skin lesions. In: DJ Manning, CK Abbey (Eds.) *Proc. Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment*, SPIE vol. 7627. Bellingham: SPIE. pp 762717-1 to 762717-10. ISBN: 9780819480286.

12. Ballerini L, Li X, Fisher RB, Aldridge B, Rees J., 2010. Content-Based Image Retrieval of Skin Lesions by Evolutionary Feature Synthesis. In: C di Chio, S Cagnoni, C Cotta, M Ebner, A Ekart, Al Esparcia-Alcazar, CK Gog, JJ Merelo, F Neri, M Preuss, J Togelius, GN Yannakakis (Eds.) Proc. Applications of Evolutionary Computation, LNCS vol. 6024. Berlin: Springer. pp 312 to 319. ISBN: 9783642122385.

13. Li X, Aldridge B, Ballerini L, Fisher B, Rees J., 2009. Depth data improves skin lesion segmentation. In: GZ Yang, D Hawkes, D Rueckert, A Noble, C Taylor (Eds.) Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), LNCS vol. 5762. Berlin: Springer. pp 1100 to 1107. ISBN: 9783642042706. PMID: 20426221

Abstracts

1. Di Leo C, Ballerini L, Fisher R, Bevilacqua V, Aldridge B, Rees J. Hierarchical Classification of Ten Skin Lesion Classes. BMVA Press. 2015.

2. Aldridge RB, Zanotto M, Ballerini L, Fisher RB, Rees JL. Using multi-dimensional scaling (MDS) and hierarchical clustering, novices can sub-classify basal cell carcinomas (BCCs) in a similar manner to experienced dermatologists. *J Investig Dermatol*, 2012; 132: 70.

3. Aldridge RB, Naysmith L, Ooi ET, Murray C, Rees J. Teledermatology triage of suspicious skin lesions potentially could be missing the majority of melanomas. *Br J Dermatol*, 2012; 167(S1): 3.

4. Aldridge R, Rees J. Improving the learning curves for undergraduate identification of skin cancer: medical students' diagnostic accuracy is raised significantly through even limited virtual exposure. *J Investig Dermatol*, 2011; 131: 52.

5. Aldridge B, Stephenson K, Naysmith L, Rees J. Current dermatological textbooks and photographic atlases do not adequately supplement students' learning, as the image content may not be representative of real-life lesions. *J Investig Dermatol*, 2011; 131: 52.

6. Aldridge RB, Xi L, Fisher RB, Rees JL. 3-Dimensional (3D) Models of Skin Lesions Hold significantly more diagnostic information and are undergraduate students' preferred adjunct for dermatology education. *Medical Education* 2011; 45(S2): 40.

7. Aldridge RB, Maxwell S, Rees J. Dermatology undergraduate clinical exposure is inadequate to meet the current British Association of Dermatologists' guidelines. *Br J Dermatol*, 2011; 165(S1): 8

8. Aldridge RB, Zanotto M, Ballerini L, Fisher RB, Rees JL. The 'ABCD' mnemonic does not function as a useful guide in assisting novices with the diagnosis of melanoma. *Br J Dermatol*, 2011; 165(S1): 55-56.

9. Aldridge RB, Maxwell S, Rees J. What exactly do students learn in their undergraduate dermatology attachment? The effects of current teaching on students' diagnostic accuracy for skin cancer and its mimics. *Br J Dermatol*, 2011; 165(S1): 60-61.

10. Maxwell S, Aldridge RB, Rees J. What resources do medical students use in order to gain additional dermatology exposure and increase their diagnostic confidence? *Br J Dermatol*, 2011; 165(S1): 135.

11. Aldridge R, Glodzik D, Bisset Y, Ballerini L, Robertson K, Xi L, Naysmith L, Fisher R, Rees J. Dermofit: a novel software that improves novices' diagnostic accuracy to a level above that of trained medical students. *J Investig Dermatol*, 2010; 130 (S72).

12. Aldridge RB, Fisher R, Ballerini L, Robertson K, Bisset Y, Rees JL. Do laypersons have intrinsic pattern recognition abilities that could be harnessed to allow the accurate and early diagnosis of skin cancers? *Br J Dermatol*, 2010; 162: 41.

International Presentations

1. AAD, March 2013 (American Academy of Dermatology), Miami
(Invited Oral- Aldridge RB) Digital simulation's role in skin cancer education: what, why, when, how, where & who? Winner of SDS Travel scholarship
2. Doctors Update, January 2013, (Combined Dermatology/Plastics Section) Val d'Isere
(Invited Oral- Aldridge RB) The identification of skin malignancies: the rules we ought (and ought not) to know.
3. TheSIS, June 2012 (The Skin Investigation Society), London
(Invited Oral- Aldridge RB) Clinical dermato-informatics: a more relevant research experience than the wet bench?
- 4.ESDR, Sept' 2012 (European Society for Dermatology Research), Venice
(Oral- Aldridge RB) Using multi-dimensional scaling (MDS) and hierarchical clustering, novices can sub-classify basal cell carcinomas (BCCs) in a similar manner to experienced dermatologists.
5. ISBI, May 2012 (International Symposium on Biomedical Imaging), Barcelona
(Oral- Ballerini L) Non-Melanoma skin lesion classification using colour image data in a hierarchical K-NN classifier.
6. ESDR, Sept' 2011 (European Society for Dermatology Research), Barcelona
(Oral- Aldridge RB) Improving the learning curves for undergraduate identification of skin cancer: medical students' diagnostic accuracy is raised significantly through even limited virtual exposure.
7. ESDR, Sept' 2011 (European Society for Dermatology Research), Barcelona
(Poster) Current dermatological textbooks and photographic atlases do not adequately supplement students' learning, as the image content may not be representative of real-life lesions.
8. ASME, July 2011 (Association for the Study of Medical Education), Edinburgh
(Oral- Aldridge RB) 3-Dimensional (3D) Models of Skin Lesions Hold significantly more diagnostic information and are undergraduate students' preferred adjunct for dermatology education.
9. ISBI, April 2011 (International Symposium on Biomedical Imaging), Chicago
(Poster) Estimating the ground truth from multiple individual segmentations incorporating prior pattern analysis with application to with application to skin lesion segmentation.
10. SPIE Medical Imaging, Feb' 2011, (Society of Photo-Optical Instrumentation Engineers), San Diego
(Oral- Zanotto M) Visual Cues Do Not Improve Lesion ABC(D) Grading.
11. ESDR, Sept' 2010 (European Society for Dermatology Research), Helsinki
(Poster) Dermofit: a novel software that improves novices' diagnostic accuracy to a level above that of trained medical students.
12. MIUA, July 2010 (Medical Image Understanding and Analysis), Warwick
(Poster) Estimating the ground truth from multiple individual segmentations with application to skin lesion segmentation.
13. EvoApplications, April 2010 (Applications of Evolutionary Computation), Istanbul
(Poster) Content-Based Image Retrieval of Skin Lesions by Evolutionary Feature Synthesis.

14. SPIE Medical Imaging, Feb' 2010 (Society of Photo-Optical Instrumentation Engineers), San Diego
(Oral- Ballerini L) Fuzzy description of skin lesions.
15. MCBR-CDS, Sept' 2009 (Medical Content Based Retrieval for Clinical Decision Support), London
(Oral- Ballerini L) A query-by-example content-based image retrieval system of non-melanoma skin lesions.
16. MICCAI, Sept' 2009 (Medical Image Computing & Computer Assisted Intervention), London
(Poster) Depth data improves skin lesion segmentation.

National Presentations

1. MIA, March 2015 (Medical Image Analysis), Dundee
(Oral- Di Leo C) Hierarchical Classification of Ten Skin Lesion Classes.
2. BAD, July 2012 (British Association of Dermatologists), Birmingham
(Oral- Aldridge RB) Teledermatology triage of suspicious skin lesions potentially could be missing the majority of melanomas.
3. SDS, May 2012 (Scottish Dermatology Society), Aberdeen
(Oral- Aldridge RB) Improving non-experts' identification of skin malignancies: virtual exposure can raise students' diagnostic accuracy significantly. Winner of SDS Prize for best presentation
4. SMG, December 2011 (Scottish Melanoma Group), Glasgow
(Oral- Naismith L) The importance of examining patients.
5. BAD, July 2011 (British Association of Dermatologists), London
(Oral- Aldridge RB) Dermatology undergraduate clinical exposure is inadequate to meet the current British Association of Dermatologists' guidelines.
6. BAD, July 2011 (British Association of Dermatologists), London
(Poster) The 'ABCD' mnemonic does not function as a useful guide in assisting novices with the diagnosis of melanoma.
7. BAD, July 2011 (British Association of Dermatologists), London
(Poster) What exactly do students learn in their undergraduate dermatology attachment? The effects of current teaching on students' diagnostic accuracy for skin cancer and its mimics.
8. BTS, July 2011 (British Teledermatology Society), London
(Oral- Aldridge RB) What resources do medical students use in order to gain additional dermatology exposure and increase their diagnostic confidence?
9. SDS, June 2011 (Scottish Dermatology Society), Edinburgh
(Oral- Maxwell S- Supervised SSM Student) Although undergraduate dermatology attachments temporarily improve students' diagnostic acumen, the clinical exposure witnessed is inadequate to allow the students to attain the BAD's recommended graduate competencies
10. BSID, April 2010 (British Society for Investigative Dermatology), Edinburgh
(Poster) Do laypersons have intrinsic pattern recognition abilities that could be harnessed to allow the accurate and early diagnosis of skin cancers?

Appendix of published work

Copies of the author's full text publications associated with his work on the Dermofit project are attached below. These are presented in chronological order but have been split into two sections: First the clinical publications encompassing the work presented in this thesis (Chapters 4, 5, 7, 9, 10). Second the collaborative informatics work outlined in the preface. The work is all open access (a Wellcome project grant requirement) and I can confirm I have the co-authors' approval to reproduce this material.

Clinical publications

1. 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 Derm Venereol.* 2013; 93(6): 689-92. PMID 23695107.
2. Aldridge RB, Maxwell SS, Rees JL. Dermatology undergraduate skin cancer training: A disconnect between recommendations, clinical exposure and competence. *BMC Medical Education* 2012; 12:27. PMID: 22569037.
3. Aldridge RB, Glodzik D, Ballerini L, Fisher RB, Rees JL. Utility of non-rule-based visual matching as a strategy to allow novices to achieve skin lesion diagnosis. *Acta Derm Venereol.* 2011; 91(3): 279-83. PMID: 21461552.
4. Aldridge RB, Zanotto M, Ballerini L, Fisher RB, Rees JL. Novice identification of melanoma: not quite as straightforward as the ABCDs. *Acta Derm Venereol.* 2011; 91(2): 125-30. PMID: 21311845.
5. Aldridge RB, Li X, Ballerini L, Fisher RB, Rees JL. Teaching dermatology using 3-dimensional virtual reality. *Arch Dermatol.* 2010; 146(10):1184-5; PMID: 20956667.

Informatics publications

6. Ballerini L, Fisher RB, Aldridge B, Rees J, 2012. A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. In: ME Celebi, G Schaefer (Eds.) *Color Medical Image Analysis*, Springer (London) pp. 63-86. ISBN 9789400753884.
7. Ballerini L, Fisher RB, Aldridge B, Rees J., 2012. Non-Melanoma skin lesion classification using colour image data in a hierarchical K-NN classifier. In: A Frangi, A Santos (Eds.) *Proc. International Symposium on Biomedical Imaging (ISBI): From Nano to Macro*, IEEE pp. 358 to 361. ISBN: 978145771857.
8. Li X, Aldridge B, Rees J, Fisher R., 2011. Estimating the ground truth from multiple individual segmentations incorporating prior pattern analysis with application to with

- application to skin lesion segmentation. In: S Wright, X Pan, M Liebling (Eds.) Proc. International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, IEEE pp. 1438 to 1441. ISBN: 9781424441273.
9. Zanutto M, Ballerini L, Fisher RB, Aldridge B, Rees J., 2011. Visual Cues Do Not Improve Lesion ABC(D) Grading. In: DJ Manning, CK Abbey (Eds.) Proc. Medical Imaging 2011: Image Perception, Observer Performance, and Technology Assessment, SPIE vol. 7966. Bellingham: SPIE. pp. 79660U-1 to 79660U-10. ISBN: 9780819485083.
 10. Li X, Aldridge B, Rees J, Fisher R., 2010. Estimating the ground truth from multiple individual segmentations with application to skin lesion segmentation. In: AH Bhalerao, NM Rajpoot (Eds.) Proc. Medical Image Understanding and Analysis (MIUA), BMVA Press. pp. 101 to 106. ISBN: 9780956615008.
 11. Laskaris N, Ballerini L, Fisher RB, Aldridge B, Rees J., 2010. Fuzzy description of skin lesions. In: DJ Manning, CK Abbey (Eds.) Proc. Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment, SPIE vol. 7627. Bellingham: SPIE. pp 762717-1 to 762717-10. ISBN: 9780819480286.
 12. Ballerini L, Li X, Fisher RB, Aldridge B, Rees J., 2010. Content-Based Image Retrieval of Skin Lesions by Evolutionary Feature Synthesis. In: C di Chio, S Cagnoni, C Cotta, M Ebner, A Ekart, AI Esparcia-Alcazar, CK Gog, JJ Merelo, F Neri, M Preuss, J Togelius, GN Yannakakis (Eds.) Proc. Applications of Evolutionary Computation, LNCS vol. 6024. Berlin: Springer. pp 312 to 319. ISBN: 9783642122385.
 13. Li X, Aldridge B, Ballerini L, Fisher B, Rees J., 2009. Depth data improves skin lesion segmentation. In: GZ Yang, D Hawkes, D Rueckert, A Noble, C Taylor (Eds.) Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), LNCS vol. 5762. Berlin: Springer. pp 1100 to 1107. ISBN: 9783642042706. PMID: 20426221