

This is a repository copy of *Predictive accuracy of risk prediction models for recurrence, metastasis and survival for early-stage cutaneous melanoma : a systematic review.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/210693/>

Version: Published Version

---

**Article:**

Kunonga, Tafadzwa Patience, Kenny, R. P.W., Astin, Margaret et al. (11 more authors) (2023) Predictive accuracy of risk prediction models for recurrence, metastasis and survival for early-stage cutaneous melanoma : a systematic review. *BMJ Open*. e073306. ISSN 2044-6055

<https://doi.org/10.1136/bmjopen-2023-073306>

---

**Reuse**


This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# BMJ Open Predictive accuracy of risk prediction models for recurrence, metastasis and survival for early-stage cutaneous melanoma: a systematic review

Tafadzwa Patience Kunonga <sup>1,2</sup>, R P W Kenny,<sup>1,2</sup> Margaret Astin,<sup>1</sup> Andrew Bryant,<sup>3</sup> Vasileios Kontogiannis,<sup>4</sup> Diarmuid Coughlan,<sup>4</sup> Catherine Richmond,<sup>1,2</sup> Claire H Eastaugh,<sup>1,2</sup> Fiona R Beyer,<sup>1,2</sup> Fiona Pearson,<sup>1,2</sup> Dawn Craig,<sup>1,2,4</sup> Penny Lovat,<sup>5,6</sup> Luke Vale,<sup>4</sup> Robert Ellis<sup>5,6,7</sup>

**To cite:** Kunonga TP, Kenny RPW, Astin M, *et al*. Predictive accuracy of risk prediction models for recurrence, metastasis and survival for early-stage cutaneous melanoma: a systematic review. *BMJ Open* 2023;**13**:e073306. doi:10.1136/bmjopen-2023-073306

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2023-073306>).

Received 02 March 2023  
Accepted 03 September 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Tafadzwa Patience Kunonga; [patience.kunonga@newcastle.ac.uk](mailto:patience.kunonga@newcastle.ac.uk)

## ABSTRACT

**Objectives** To identify prognostic models for melanoma survival, recurrence and metastasis among American Joint Committee on Cancer stage I and II patients postsurgery; and evaluate model performance, including overall survival (OS) prediction.

**Design** Systematic review and narrative synthesis.

**Data sources** Searched MEDLINE, Embase, CINAHL, Cochrane Library, Science Citation Index and grey literature sources including cancer and guideline websites from 2000 to September 2021.

**Eligibility criteria** Included studies on risk prediction models for stage I and II melanoma in adults  $\geq 18$  years. Outcomes included OS, recurrence, metastases and model performance. No language or country of publication restrictions were applied.

**Data extraction and synthesis** Two pairs of reviewers independently screened studies, extracted data and assessed the risk of bias using the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies checklist and the Prediction study Risk of Bias Assessment Tool. Heterogeneous predictors prevented statistical synthesis.

**Results** From 28 967 records, 15 studies reporting 20 models were included; 8 (stage I), 2 (stage II), 7 (stages I–II) and 7 (stages not reported), but were clearly applicable to early stages. Clinicopathological predictors per model ranged from 3–10. The most common were: ulceration, Breslow thickness/depth, sociodemographic status and site. Where reported, discriminatory values were  $\geq 0.7$ . Calibration measures showed good matches between predicted and observed rates. None of the studies assessed clinical usefulness of the models. Risk of bias was high in eight models, unclear in nine and low in three. Seven models were internally and externally cross-validated, six models were externally validated and eight models were internally validated.

**Conclusions** All models are effective in their predictive performance, however the low quality of the evidence raises concern as to whether current follow-up recommendations following surgical treatment is

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Comprehensive and systematic searches captured the core evidence about prediction models in early-stage melanoma.
- ⇒ The current evidence reviewed across all models is insufficient to make recommendations regarding use in clinical practice.
- ⇒ Heterogeneity in the data across study populations, predictors and clinical progression of the disease suggests there is insufficient evidence to make firm conclusions about best clinical practice in the application of predictive models in patients with American Joint Committee on Cancer stage I or II cutaneous melanoma.

adequate. Future models should incorporate biomarkers for improved accuracy.

**PROSPERO registration number** CRD42018086784.

## INTRODUCTION

Melanoma is one of the deadliest of all skin cancers.<sup>1</sup> The worldwide incidence of melanoma is estimated to be approximately 2% of cancers per annum.<sup>2</sup> However, in early-stage disease, once surgically removed from the skin, through wide local excision (WLE), and without adjuvant immunotherapy, chemotherapy or radiotherapy, the majority of early-stage melanomas are cured, with an estimated 5-year overall survival (OS) rate of 95%–100% (stage I) and 65%–92.8% (stage II).<sup>3</sup> However, up to 30% of all primary melanomas progress to metastatic disease with an associated extremely poor survival rate of only 5%–15%.<sup>4</sup> Although the surgical treatment of primary melanoma is effective and long established, there has been a rapid pace of change recently with the addition of sentinel lymph node biopsy (SLNB).<sup>5</sup> This

procedure identifies the sentinel lymph node which is excised for examination to determine whether cancer cells are present.<sup>6</sup>

Currently, a structured, uniformly adopted, evidence-based model of patient follow-up after initial diagnosis and treatment is lacking.<sup>7</sup> Follow-up strategies depend on the stage and progression of the disease; based on the criteria defined within the eighth edition of the American Joint Committee on Cancer (AJCC) staging manual.<sup>8</sup> For this review, however, the definitions of stage and progression used in the 2008, seventh edition<sup>9</sup> and earlier editions remain relevant as the reported data in research studies may be based on definitions from earlier editions. Current guidelines for management of the condition vary across the world, with most developed using anecdotal evidence and expert opinion. A clinical prediction model, a statistical tool that relates multiple factors to the probability of a patient having a future clinical event,<sup>10</sup> could be used to aid early detection of disease progression such as new primary tumours, in-transit metastasis or locoregional recurrence in the regional lymph nodes. Before introducing a new clinical prediction model into practice there needs to be evidence of model development, validation and impact. Validation should involve evaluating performance (in terms of discrimination or calibration), and clinical usefulness in practice. Impact should consider whether the model improves decision-making.<sup>10</sup>

Previous systematic reviews have been conducted looking at the value of risk prediction models for melanoma development,<sup>11–13</sup> and clinical risk of survival of melanoma.<sup>14</sup> However, to our knowledge, no review has critically appraised current clinicopathological prediction models for primary melanomas (stages I and II), following surgical excision of the tumour. This review aims to identify and assess prognostic/predictive models used to predict patient survival, recurrence (any site) and metastasis in AJCC stage (I and II) melanoma following surgical excision, thus potentially allowing further refinement of risk-stratification of patients.

## METHODS

The review population listed in the protocol was later augmented to include stage II melanoma. As with stage I melanoma, tumours at this stage may not have advanced and are treated with WLE.<sup>15</sup> The review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guideline (PRISMA).<sup>16</sup>

### Search strategy

Searches were originally conducted in July 2019, then updated in September 2021. A date restriction of 2000 was used, as this is when SLNB became routine in the USA.<sup>5</sup> The search strategy (online supplemental tables 1 and 2) was designed in MEDLINE and translated to other databases (online supplemental table 3). A published and validated prognostic study filter<sup>17</sup> was used including the following concepts: (melanoma) AND (risk models)

AND (prognosis). No language or country of publication restrictions were applied. In the search update (June 2019 to September 2021), a more up to date search filter was used.<sup>18</sup>

### Inclusion and exclusion criteria

Studies were selected using piloted inclusion criteria (table 1) according to the Patient population, Intervention, Comparator, Outcome, Timing and Setting formula.<sup>19</sup> The primary outcome is OS defined as patient survival until death from melanoma following primary treatment. Secondary outcomes include: number of detected recurrence and metastasis, and the discrimination, calibration, overall performance, and clinical utility of the models. A table of definitions for the performance measures is presented in Online supplemental table 4.

### Selection of studies

Two reviewers independently screened titles and abstracts for articles relevant to stage I disease. For stage II screening a text mining approach was adopted, followed by duplicate and independent hand screening of study abstracts. Two reviewers independently screened the selected full text articles. For each included study, data were extracted by one of four reviewers using the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) checklist.<sup>20</sup> Data were extracted for characteristics of participants, outcomes, predictors, model development methods, model performance and validation. Extraction forms were checked for accuracy and consistency by another reviewer, and a clinical expert.

### Risk of bias assessment

Two pairs of reviewers used the Prediction study Risk of Bias Assessment Tool (PROBAST),<sup>21</sup> to assess the risk of bias (ROB) of each study. The tool, which has gained popularity on reviews of prognostic studies,<sup>22</sup> evaluates 20 questions in four domains (Participants, Predictors, Outcome and Analysis) to assess the ROB and applicability of prediction model studies.<sup>21</sup> Following the guide, we used the signalling questions across these domains to make the following judgements: 'yes' means low bias, 'no' means high bias. For the overall judgement, the study was rated as low ROB if all domains were rated as low, high ROB if at least one domain was rated as high. Where there was insufficient information to make a judgement, we rated the domain as unclear ROB. However, if one domain was rated as unclear and the others low, overall judgement was rated as unclear.<sup>21</sup>

### Data analysis and synthesis

A narrative synthesis was undertaken, including summarising the characteristics of the included models and the performance of prediction models, focusing on measures of discrimination, calibration and overall performance; and model validation methods. We were unable to conduct a meta-analysis due to the variance in use of disparate predictors, and modelling methods.

**Table 1** Inclusion criteria

Criteria	Include	Exclude
Population	Adults aged $\geq 18$ years treated for American Joint Committee on Cancer (seventh edition) stage I and II cutaneous melanoma <ul style="list-style-type: none"> <li>▶ Stage IA (T1a <math>\leq 1</math> mm thick); or</li> <li>▶ Stage IB (T1b with ulceration or mitoses <math>\leq 1</math> mm thick, or T2a 1.01 to 2.00 mm thick and no ulceration); or</li> <li>▶ Stage IIA (1.01–2.00 mm thick with ulceration, or 2.01–4.00 mm thick without ulceration); or</li> <li>▶ stage IIB (2.01–4.00 mm thick with ulceration, or <math>&gt;4.0</math> mm without ulceration)</li> <li>▶ stage IIC (<math>&gt;4.0</math> mm with ulceration)</li> </ul>	Advanced melanoma Metastatic melanoma Stage III Stage IV
Types of prognostic models	<ul style="list-style-type: none"> <li>▶ Nomogram</li> <li>▶ Scoring system</li> <li>▶ Equation</li> <li>▶ Classification or decision trees</li> </ul> *Minimum of two clinicopathological factors	Models built with predictive biomarker factors included Models built with gene expression profiling Diagnostic studies Studies assessing only one predictor
Outcome measures	Primary: overall survival Secondary: recurrence (any site), metastases, prognostic performance of risk models	
Timing	Postresection of the primary cutaneous tumour <ul style="list-style-type: none"> <li>▶ Diagnostic excision</li> <li>▶ Re-excision, also called wide local excision (WLE)</li> </ul>	Studies that are looking at treatment (eg, therapy) of melanoma
Setting	Any setting (primary, secondary or tertiary care)	
Study design	Retrospective cohort studies Prospective cohort studies Randomised controlled trials	Non-empirical studies

Furthermore, it was not possible to perform subgroup analyses as the studies identified did not report outcomes by population subgroups.

### Quality of the evidence

We had planned to assess the overall quality of evidence using the Grading of Recommendations Assessment, Development and Evaluation.<sup>23</sup> However, although the tool has been adapted for assessing overall quality in prognostic factor reviews, studies, it is yet to be adapted for prediction model reviews.<sup>24</sup>

### Patient and public involvement

This review was requested by our funder within a timescale that did not allow for meaningful public and patient involvement.

## RESULTS

### Selection of studies

We identified 28 967 records after deduplication, of which 165 were selected for full text screening (online supplemental figure 1). Fourteen studies reporting twenty unique risk prediction models, met the inclusion criteria.<sup>25–38</sup>

### Characteristics of studies

Characteristics of included studies are presented in [table 2](#). Studies were published between 2000 and 2021; nine from the USA,<sup>26 27 30 31 33–35 37 38</sup> two from Australia,<sup>25 28</sup> two from Europe<sup>32 36</sup> and one from Brazil.<sup>29</sup> Twelve studies were from retrospective cohorts<sup>25–32 36–38</sup> and three were from prospective cohorts.<sup>33–35</sup> Patient data were obtained from population-based cancer registries, medical/clinical records or a combination of these sources. Study intervals ranged from 1972 to 2015. Eight models were developed from patients with stage I melanoma only (three from one study,<sup>28</sup> two from another study<sup>30</sup> and three from three studies<sup>31–33</sup>), two for stage II only (both reported in the same study<sup>29</sup>), eight for stages I–II (two reported in one study<sup>36</sup> and five from five studies<sup>25 26 34 35 38</sup>) and for three models the stages not reported (two from one study<sup>27</sup> and the other from one study<sup>37</sup>), however they were found to be applicable to stages I–II. Eleven studies used logistic regression methods to develop the models,<sup>25–29 32–34 36–38</sup> and three used a recursive partitioning method.<sup>30 31 35</sup> Duration of follow-up ranged from 3<sup>38</sup> to 20 years.<sup>26</sup>

### Characteristics of included models

Characteristics of included models are presented in [table 3](#). Model outcomes included OS<sup>25 27 30 32 34 35 38</sup>; melanoma specific survival (MSS)<sup>25 26 29</sup>; melanoma specific

**Table 2** Characteristics of included studies

Citation	Study location	Study design	Data source (centres (n))	AJCC disease stage/ clinical staging	Study interval	Development method
Baade <i>et al</i> <sup>25</sup>	Australia	Retrospective	Population based Queensland cancer registry	2002 AJCC 94% Stages I–II	1995–2008	Multivariate analysis
Balch <i>et al</i> <sup>26</sup>	USA	Retrospective	Population based AJCC database (13)	Stages I–II	Not reported	Multivariate analysis
Cochran <i>et al</i> <sup>27</sup>	USA	Retrospective	John Wayne Cancer Institute Melanoma clinical database, Division of Surgical Oncology, UCLA	1988 AJCC Not reported	1980–1990	Multivariate analysis
El Sharouni <i>et al</i> <sup>28</sup>	Netherlands Australia	Retrospective	DEV: PALGA, the Dutch Pathology Registry VALID: Melanoma Institute Australia	2009 AJCC Stages IA–IB IA (T1a)=58.6% IB (T1b)=29% T1nos=12.4%	2000–2014	Multivariate analysis
Fonseca <i>et al</i> <sup>29</sup>	Brazil	Retrospective	C. Camargo Cancer Centre database	2009 AJCC Stages IIB–IIC	2000–2014	Multivariate analysis
Gimotty <i>et al</i> <sup>31</sup>	USA	Retrospective	DEV: SEER population based registry VALID: New SEER patients	2002 AJCC Stages IA–IB IA (T1a) = 86% IB (T1b) = 14%	DEV: 1972– 1991 VALID: 1991–1995	Recursive partitioning
Gimotty <i>et al</i> <sup>30</sup>	USA	Retrospective	DEV: SEER Registry VALID: Clinical based PLC registry	2002 AJCC Stages IA (T1a)–IB (T1b)	DEV: 1998– 2001 VALID: 1972– 2001	Recursive partitioning
Maurichi <i>et al</i> <sup>32</sup>	Europe	Retrospective	European clinical based centres (6)	2009 AJCC Stages IA–IB T1a=50.3% T1b=49.7%	1996–2004	Multivariate analysis
Rosenbaum <i>et al</i> <sup>33</sup>	USA	Prospective	NYU clinicopathological biospecimen database	2009 AJCC Stage IB	2002–2014	Multivariate analysis
Soong <i>et al</i> <sup>34</sup>	USA	Prospective	DEV: 2008 AJCC population based Melanoma Database (9) VALID: Sydney Melanoma Unit, Australia	2008 AJCC Stages I–II	DEV: 26% diagnosed after 2002 VALID: Not reported	Multivariate analysis
Tsai <i>et al</i> <sup>35</sup>	USA	Prospective	AJCC Melanoma population-based database (13)	2002 AJCC Stages IA–IIC	Not reported	Recursive partitioning
Verver <i>et al</i> <sup>36</sup>	Europe	Retrospective	EORTC	2009 AJCC Stages IA–IIC	1997–2013	Multivariate analysis
Vollmer and Seigler <sup>37</sup>	USA	Retrospective	University Melanoma Clinic database	1988 AJCC Not reported	1980–1990	Multivariate analysis
Xiao <i>et al</i> <sup>38</sup>	USA	Retrospective	SEER Registry	2009 AJCC Stages IA–IIC	2010–2015	Multivariate analysis

AJCC, American Joint Committee on Cancer; DEV, development set; EORTC, European Organisation for Research and Treatment of Cancer Centres; NYU, New York University; RND, regional node dissection; SEER, Surveillance, Epidemiology and End Results Programme; UCLA, University of California, Los Angeles; VALID, validation set.

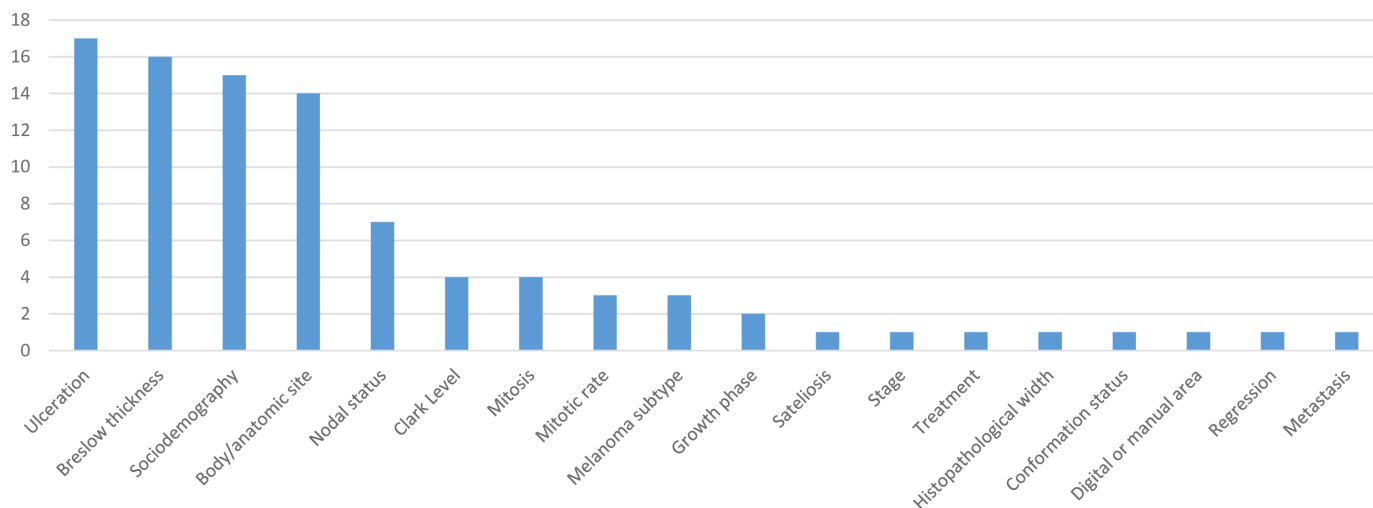
mortality (MSM)<sup>36</sup>; mortality<sup>37</sup>; recurrence<sup>27 36</sup>; recurrence free survival (RFS)<sup>33</sup>; local RFS (LRFS)<sup>28</sup>; regional RFS (RRFS)<sup>28</sup>; distant RFS (DRFS)<sup>28 29</sup> and metastasis.<sup>31</sup> A table of definitions used in the studies for model outcomes is presented in Online supplemental table 5. Total sample sizes across the models ranged from 259<sup>29</sup> to 66 192 patients.<sup>38</sup> Five studies selected candidate clinicopathological variables based on previous clinical knowledge or literature (pre-specification),<sup>25 28 30 35 37</sup> three conducted a multivariate analysis,<sup>30 32 34</sup> two studies conducted a univariate analysis,<sup>29 33</sup> while three employed

a combination of previous knowledge and univariate analysis.<sup>27 36 38</sup> One study did not provide information on variable selection.<sup>26</sup> A backward elimination analysis approach was used in seven studies,<sup>25 27–29 31 33 36</sup> and a full model approach was used in six studies,<sup>30 32 34 35 37 38</sup> to select the final predictors for model development. The number of predictors in the final models ranged from 3<sup>33 36</sup> to 10.<sup>38</sup> The most common predictors were ulceration (17/20), Breslow thickness/depth (16/20), socio-demographic status (15/20) and body/anatomic site (14/20) (see figure 1). The models are reported based on

**Table 3** Model development

Citation	Sample size	Follow-up period	End point	Events (n)	Candidate selection method	Final predictor selection	Final predictors (n)	Validation method
Baade <i>et al</i> <sup>25</sup>	28 654	Median=7.2 years (86.4 months)	MSS	1700	Prespecification from existing data	Backward elimination	7	Internal-external cross validation
Balch <i>et al</i> <sup>26</sup>	13 581	5–20 years	MSS	NR	Unclear	Unclear	8	External—geographic
Cochran <i>et al</i> <sup>27</sup>	1042	Median=42.5 months	OS	NR	Prespecification+univariate analysis	Backward elimination	5	Internal sample validation
Cochran <i>et al</i> <sup>27</sup>	1042	Median=42.5 months	Recurrence	NR	Prespecification+univariate analysis	Backward elimination	4	Internal sample validation
El Sharouni <i>et al</i> <sup>28</sup>	DEV: 25 930 VALID: 2968	DEV: Median=6.7 years VALID: Median=12 years	LRFS	232	Prespecification from existing data	Backward elimination	6	Internal-external cross validation
El Sharouni <i>et al</i> <sup>28</sup>	DEV: 25 930 VALID: 2968	DEV: Median=6.7 years VALID: Median=12 years	RRFS	564	Prespecification from existing data	Backward elimination	7	Internal-external cross validation
El Sharouni <i>et al</i> <sup>28</sup>	DEV: 25 930 VALID: 2968	DEV: Median=6.7 years VALID: Median=12 years	DRFS	278	Prespecification from existing data	Backward elimination	6	Internal-external cross validation
Fonseca <i>et al</i> <sup>29</sup>	259	Median=80.13 months	DRFS	117	Univariate analysis	Backward elimination	5	Bootstrap
Fonseca <i>et al</i> <sup>29</sup>	259	Median=80.13 months	MSS	117	Univariate analysis	Backward elimination	5	Bootstrap
Gimotty <i>et al</i> <sup>31</sup>	DEV: 884 VALID: 114	At least 10 years	Metastasis	127	Multivariate analysis	Backward elimination	4	External—new cohort
Gimotty <i>et al</i> <sup>30</sup>	DEV: 26 114	Median=4.6 years	OS	3593	Prespecification from existing data	Full model approach	4	External—geographic
Gimotty <i>et al</i> <sup>30</sup>	VALID: 2389	Median=8.1 years	OS	1076	Prespecification from existing data	Full model approach	5	External—geographic
Maurichi <i>et al</i> <sup>32</sup>	2243	124 months	12 year OS	240	Multivariate analysis	Full model approach	6	Congruence examination
Rosenbaum <i>et al</i> <sup>33</sup>	DEV: 506 VALID: 149	Median=4.4 years	RFS	NR	Univariate analysis	Backward elimination	3	Internal k-fold validation
Soong <i>et al</i> <sup>34</sup>	DEV: 14 760 VALID: 10 974	5 and 10 years	OS	NR	Multivariate analysis	Full model approach	6	External—geographic
Tsai <i>et al</i> <sup>35</sup>	13 268	Not reported	OS	NR	Prespecification from existing data	Full model approach	6	Internal-external cross validation
Verver <i>et al</i> <sup>36</sup>	3180	Median=61 months	Composite recurrence and/or MSM	Recurrence=496 Dead=277	Prespecification+univariate analysis	Backward elimination	3	Internal-external cross validation
Vollmer and Seigler <sup>37</sup>	1910	Median=7.6 years	Mortality	NR	Prespecification	Full model approach	5	Internal k-fold validation
Xiao <i>et al</i> <sup>38</sup>	DEV: 46 336 VALID: 19 856	3 and 5 years	OS	NR	Prespecification+univariate analysis	Full model approach	10	Internal sample validation

DEV, development set; DRFS, distant recurrence-free survival; EPV, events per variable; MSM, melanoma specific mortality; MSS, melanoma specific survival; NR, not reported; OS, overall survival; RFS, recurrence free survival; RRFS, regional recurrence free survival; VALID, validation set.



**Figure 1** Identified predictors.

first authors and outcome as follows: Baade MSS model<sup>25</sup>; Balch MSS model<sup>26</sup>; Cochran OS<sup>27</sup>; Cochran recurrence risk score<sup>27</sup>; El Sharouni LRFS nomogram<sup>28</sup>; El Sharouni, RRFs nomogram<sup>28</sup>; El Sharouni DRFS nomogram<sup>28</sup>; Fonseca, DRFS nomogram<sup>29</sup>; Fonseca MSS nomogram<sup>29</sup>; Gimotty metastasis nomogram<sup>31</sup>; Gimotty(a) OS model<sup>30</sup>; Gimotty(b) OS model<sup>30</sup>; Maurichi OS model<sup>32</sup>; Rosenbaum RFS model<sup>33</sup>; Soong OS model<sup>34</sup>; Tsai OS model<sup>35</sup>; Verver MSM model<sup>36</sup>; Vollmer mortality model<sup>37</sup> and Xiao OS model.<sup>38</sup>

## Performance of prediction models

### Discrimination

The Area Under the Curve (AUC) - Receiver Operating Characteristics (ROC), also known as AUC-ROC was reported in two models (reported in one study) predicting OS,<sup>30</sup> one model predicting RFS<sup>33</sup> and another predicting metastasis outcomes,<sup>31</sup> all with values above 0.7. The C-index measure was reported in the Maurichi OS model<sup>32</sup>; Xiao OS model<sup>38</sup>; El Sharouni DRFS model<sup>28</sup>; Fonseca DRFS model<sup>29</sup>; El Sharouni LRFS model<sup>28</sup>; El Sharouni RRFs model<sup>28</sup>; Fonseca MSS model<sup>29</sup> and Verver MSM model,<sup>36</sup> all with values above 0.7 for the models. The discriminative performance of the models is considered acceptable when the AUC-ROC statistics and their equivalent are  $\geq 0.7$ .<sup>39</sup> In addition to the C-index, the Baade MSS model also estimated the D-statistic and reported a value of 1.50 (95% CI 1.44 to 1.56) for OS.<sup>25</sup>

### Calibration

Calibration scores were reported in the Soong OS model,<sup>34</sup> with correlation coefficients of 0.90 and 0.93 for 5-year and 10-year survival rates.<sup>34</sup> Calibration plots for predicting survival outcomes were presented in 10 models: Maurichi OS model<sup>32</sup>; Xiao OS model<sup>38</sup>; El Sharouni DRFS model<sup>28</sup>; Fonseca DRFS model<sup>29</sup>; El Sharouni RRFs model<sup>28</sup>; El Sharouni LRFS model<sup>28</sup>; Fonseca MSS model<sup>29</sup>; Verver recurrence model<sup>36</sup> and Verver MSM model.<sup>36</sup> Three models that reported calibration scores reported values closer to 1, indicating

a perfect agreement, while eight models graphically presented plots showing a good match between predicted and observed outcomes for OS, DRFS, RRFs, LRFS, MSS and metastasis. This suggests that the models have the ability to accurately generate predictions that are close to the observed outcomes. Results indicated high accuracy and precision of the prediction models, as predicted and actual survival probabilities were similar.<sup>40</sup>

### Overall performance

The Baade OS model assessed the overall performance of predicting OS by assessing how well the model fit the data using the r-squared ( $R^2$ ) statistic.<sup>25</sup> The model reported the  $R^2$  as 0.47 (0.45 to 0.49), equating to a strong fit.<sup>25</sup> The Tsai OS model,<sup>35</sup> measured overall performance by assessing the Brier score, with usual values ranging from 0 (total accuracy) to 1 (wholly inaccurate).<sup>41</sup> Results show that the score for the proposed, prognostic classification scheme increased with time from 0.02 at year 1 to approximately 0.20 at year 15, indicating high accuracy in predicting OS.<sup>35</sup>

### Clinical utility

None of the studies conducted either a net benefit or decision curve analysis, therefore we were unable to address the clinical utility of the tools. The clinical utility of a validated risk prediction model would help clinicians and patients during the surveillance phase of care.<sup>42</sup>

### Model validation methods

Model validation methods are presented in Online supplemental table 6. Internal validation refers to the performance of a model developed and tested on a sample of the same population.<sup>43</sup> External validation refers to how well a model predicts an outcome in a data set that is different from the development population (new cohort), or a different region or country (geographical).<sup>44</sup> Eight models were internally validated. The Rosenbaum RFS model,<sup>33</sup> and Vollmer mortality model,<sup>37</sup> used the k-fold cross-validation procedure, where an original sample is

randomly partitioned into  $k$  equal-sized subsamples.<sup>45</sup> The Cochran OS model,<sup>27</sup> Cochran recurrence model<sup>27</sup> and Xiao OS model,<sup>38</sup> randomly selected a proportion of the sample and retained these as the validation set, using the rest for training. The Maurichi OS model<sup>32</sup> was validated by assessing the congruence of expected outcomes (predicted from the model) and observed outcomes. The Fonseca DRFS model,<sup>29</sup> and Fonseca MSS model,<sup>29</sup> were internally validated using bootstrap methods. Bootstrapping entails repeatedly drawing random samples with replacement from the data to construct prediction models and evaluating model performance using out-of-bag observations.<sup>41</sup> Five models validated their models using external techniques, including geographic validation: the Gimotty(a) OS model<sup>30</sup>; Gimotty(b) OS model<sup>30</sup>; Balch MSS model<sup>26</sup>; Soong OS model<sup>34</sup> and Gimotty metastasis model.<sup>31</sup> Seven models: El Sharouni LRFS model<sup>28</sup>; El Sharouni RRFS model<sup>28</sup>; El Sharouni DRFS model<sup>28</sup>; Verver recurrence model<sup>36</sup>; Verver MSM model<sup>36</sup>; Baade MSS model<sup>25</sup> and the Tsai OS model,<sup>35</sup> assessed model performance using internal-external cross-validation methods (IECV). IECV analyses the performance of models using new patients from different but related practices as compared with the original development sample.<sup>41</sup>

### Critical appraisal of included studies

The results of the critical appraisal are presented in Online supplemental table 7. Overall, eight models: Cochran OS model<sup>27</sup>; Cochran recurrence model<sup>27</sup>; Fonseca DRFS model<sup>29</sup>; Fonseca MSS model<sup>29</sup>; Verver recurrence model<sup>36</sup>; Verver MSM model<sup>36</sup>; Rosenbaum RFS model<sup>33</sup> and the Xiao OS model,<sup>38</sup> were judged to be at high risk of bias. Three models: Maurichi OS model<sup>32</sup>; Soong OS model<sup>34</sup> and Xiao OS model,<sup>38</sup> were rated as low risk. Nine models: El Sharouni LRFS model<sup>28</sup>; El Sharouni RRFS model<sup>28</sup>; El Sharouni DRFS model<sup>28</sup>; Gimotty(a) OS model<sup>30</sup>; Gimotty(b) model<sup>30</sup>; Baade OS model<sup>25</sup>; Balch OS model<sup>26</sup>; Gimotty metastasis model<sup>30</sup> and Vollmer mortality model.<sup>37</sup> The most notable cause of high risk was for the analysis domain. This was mainly due to the inclusion of variables in the final model, previously identified as significant following a univariate analysis.<sup>27 29 33 38</sup> This method can lead to incorrect predictor selection because predictors are chosen on the basis of their statistical significance as a single predictor rather than in context with other predictors.<sup>41</sup>

The applicability ratings of the models are presented in Online supplemental table 8. Two models: Cochran OS model<sup>27</sup> and Cochran recurrence model, were rated as having a high risk regarding applicability, the rest of the models were rated as low risk of bias regarding applicability. Detailed information on critical appraisal is available in Online supplemental file 1. The main concern regarding applicability was for the selection of participants in the study by Cochran *et al*,<sup>27</sup> where little information is reported about the characteristics of the patients or the severity of their melanoma.

### DISCUSSION

This review identified 14 studies describing 20 different models developed for the prediction of recurrence, new primary tumours or metastasis in patients with AJCC stage I or II cutaneous melanoma following excision. The models differed in the predictors used depending on the outcome of interest and statistical measures used to assess model performance. It was therefore inappropriate to statistically synthesise their results. One of the limitations of the studies was the lack of reporting of baseline rates of SLNB, a technique useful for providing disease stage and to guide adjuvant systemic therapy.<sup>46</sup> The absence of these rates makes it difficult to be sure that the patients are correctly classified in the studies. As the AJCC classification criteria have changed over time, studies from different periods used different staging systems and methodologies. For models at stages I or II we considered which AJCC staging criteria would apply taking into account reported factors such as tumour thickness, and ulceration (see table 1). However, this can lead to significant heterogeneity in the data, making it difficult to compare and synthesise findings. Common risk factors included in the models were consistent with well-established risk factors for melanoma, including ulceration, age, sex, Breslow thickness and tumour site.<sup>47</sup> Although systematic reviews of prediction models for cutaneous melanoma<sup>11 12</sup> have previously been conducted, to the best of our knowledge, this review investigates the potential of clinical models to predict recurrence, metastases and survival of AJCC stage I and II melanoma following surgical excision. Similar to the reviews by Vuong *et al*,<sup>12</sup> and Mahar *et al*,<sup>14</sup> notable differences in the approaches used to select predictors during model development were observed. Eight models selected predictors based on univariate analysis of the strength of these predictors, of which four were developed based on previous knowledge. When developing models, building on previous literature and expert clinician opinions is recommended. Building models based only on the statistical significance of the association between predictor and outcome in univariate analyses risks missing important predictors.<sup>48</sup>

Model performance measures were available for assessing discrimination in 14 models, calibration in 11 models and overall performance in 2 models. While the type of discrimination statistic varied, the discrimination statistics of the new models ranged from 0.72 to 0.88, a range comparable to that reported in other published melanoma reviews: 0.62 to 0.86,<sup>12</sup> 0.7 to 0.8<sup>11</sup> and 0.62 to 0.98.<sup>14</sup> Currently, there is a lack of evidence for the clinical performance of early stage melanoma prediction models based on clinicopathological variables. Accuracy measures such as calibration and discrimination, however, do not provide any detail regarding a model's suitability for clinical practice. Net benefit, obtained via decision curve analysis, aims to assess the clinical utility of such models.<sup>42</sup> Unfortunately, the studies included in this review do not give detail on clinical utility for the models they report on. In addition, we reviewed follow-up schedules as recommended by clinical guidelines and note the limited evidence



base on which they are based.<sup>49</sup> This suggests more detailed studies about timing and frequency of relapse are needed. Most of the models were externally validated, either in new cohorts or patients from a different location or through IECV. External validation of models is essential to support the general applicability of any prediction model.<sup>48</sup> However, most of the models were rated as being at high risk of bias implying low confidence in the performance of these models in new datasets. This is because a high risk of bias can lead to either overestimation or underestimation of the predictive accuracy of a model, which can affect its generalisability to new datasets.<sup>44</sup>

This review followed procedures documented by the Cochrane Collaboration for conducting systematic reviews, the CHARMS<sup>20</sup> guidelines for extracting data, the PRISMA<sup>16</sup> guidelines for reporting and PROBAST tool<sup>21</sup> for assessing risk of bias, so was therefore robust. The update search conducted in September 2021 incorporated emerging trial evidence, so we are confident that we would have identified relevant trials/studies coming to light in the next 1–2 years from the search implementation. Although the PROBAST tool provides a structured framework for evaluating various aspects of study methodology, like any other tool, it has its limitations. From our limited experience of the tool, we found that its effectiveness relied heavily on the accuracy and comprehensiveness of information available in the included studies. Most of the studies incorporated within our analysis were conducted before guidelines for conducting prognostic studies were developed, leading to missing or inconsistent information, thereby impeding a comprehensive application of the PROBAST tool to adequately appraise bias. A critical analysis of the PROBAST tool was critical of the inter-rater reliability (IRR) of the tool before training.<sup>50</sup> Although we did not calculate the IRR, we assessed RoB in two pairs to ensure consistent and reliable assessments.

A recent guideline has been developed in 2015, the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD), to ensure comprehensive reporting of critical methodological elements in individual prognostic studies.<sup>51</sup> Nevertheless, it is imperative to acknowledge that a limitation of the current review stems from the fact that the included studies were published prior to the availability of this reporting guidance. Recent evaluation of study quality based on the TRIPOD statement reveals insufficient reporting levels for risk prediction models in cutaneous melanoma. Embracing reporting guidelines such as TRIPOD, can enhance future research reporting standards.<sup>52</sup> Second, none of the studies reported a decision curve analysis to assess the clinical implications of the models. Third, comparing the included models was problematic and meta-analysis was not feasible due to the variety in the predictors and statistical measures used for model performances. Finally, we were unable to access the development and validation studies of some of the publicly available online tools, and these were excluded from our analysis.<sup>53 54</sup>

The results of this systematic review highlight the relative lack of appropriate evidence underpinning current melanoma prediction tools to support practice in AJCC stage

I and II disease. The evidence needed for clinical guideline decision-makers to incorporate a prediction model into routine clinical care needs to be far more convincing than what has been reported thus far. The goal should be clear demonstration of external validity, by evaluating model transportability in other cohorts,<sup>10</sup> and clinical utility (the ability to better select patients from high-risk groups for adjuvant therapy in the future) by undertaking decision curve analysis to identify net benefits.<sup>42 55</sup> Additionally, biomarkers play a crucial role in risk prediction models of cutaneous melanoma.<sup>56</sup> These may include genetic, immune or circulatory biomarkers.<sup>56</sup> They are measurable characteristics that provide information about the disease's presence, progression and response to treatment.<sup>56</sup> The relationship between biomarkers and clinicopathological characteristics has been shown to be interconnected in predicting chemotherapy response in breast cancer.<sup>57</sup> Therefore, incorporating biomarkers into risk prediction models alongside clinicopathological factors, may provide additional information to refine risk prediction models in cutaneous melanoma and aid healthcare professionals obtain more comprehensive and accurate assessments of an individual's risk. Additionally, other studies have shown that mutation status is associated with survival outcomes,<sup>58 59</sup> however we did not evaluate these studies as they were not developed into risk models. Critically, these results clearly outline the need for ongoing prognostication studies and, as such, this review acts as an evidence base to catalyse project development and funding, and future follow-up guidelines and management of patients, given the relative scarcity of evidence-based practice at present.

## CONCLUSION

The data, which are heterogeneous in terms of biology and progression, do not offer a wide enough scope of best practice to allow accurate prognostication of melanoma AJCC stages I or II patients as defined by the criteria within the third, sixth, seventh and eighth editions of AJCC staging, or recommendation for use in clinical practice. This raises concern as to whether current follow-up recommendations postsurgical treatment is adequate as the evidence supporting such recommendation is sparse. Future research should focus on validating existing models utilising TRIPOD guidelines to improve reporting quality. Future studies should also look to use decision curve analysis to analyse the net benefit of using the predictive model.<sup>55 60</sup>

### Author affiliations

<sup>1</sup>Evidence Synthesis Group, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

<sup>2</sup>NiHR Innovation Observatory, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

<sup>3</sup>BioStatistics Research Group, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

<sup>4</sup>Health Economics Group, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

<sup>5</sup>Dermatological Sciences, Translation and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK

<sup>6</sup>AMLo Biosciences, The Biosphere, Newcastle Helix, Newcastle upon Tyne, UK

<sup>7</sup>Department of Dermatology, South Tees Hospitals NHS FT, Middlesbrough, UK

**Acknowledgements** We would like to thank our colleagues from the Evidence Synthesis Group, and NIHR Innovation Observatory, of Newcastle University and Dr Batoul Nasr from Dermatological Sciences, Translational Research Institute, Newcastle University for their support in screening.

**Collaborators** N/A.

**Contributors** TPK was responsible for the conceptualisation of this project, methodology, writing of the original draft, visualisation of results, the overall project administration and is the guarantor of this work. RE codesigned the protocol, undertook all stages of the review and cowrote the paper. RPWK, MA, VK and DCoughlan undertook all stages of the review and cowrote the paper. AB codesigned the protocol undertook all stages of the review and cowrote the paper. CR and CHE undertook the searches for the review and cowrote the paper. FB, DCraig, PI and LV codesigned the protocol and cowrote the paper. FP contributed to the analysis of review data and cowrote the paper.

**Funding** This work was supported by the National Institute for Health Research (NIHR) Health Technology Assessment Programme, grant number 16/166/05 and the NIHR Invention for Innovation (i4i) Innovative Prognostic Test for Early-Stage Cutaneous Melanoma, grant number 20993.

**Competing interests** TPK, RPWK, MA, AB, VK, DCoughlan, CR, CHE, FB and FP have no conflicts of interest to declare. DCraig: December 2018 to present, member of the HD&DR Research-led prioritisation committee. LV was a member of the NIHR Clinical Evaluation and Trial Panel from 2015 to 2018. LV and DCraig are both part funded by the NIHR Applied Research Collaboration for the North East and North Cumbria. PL is Chief Scientific Officer for AMLo Biosciences, holds shares in AMLo and is named as an inventor on patents for biomarkers this area. RE received personal fees from AMLo Biosciences, outside the submitted work, he also was previously Chief Medical Officer for AMLo Biosciences but gave up this position in January 2021. He is named as an inventor on patents owned by AMLo Biosciences for biomarkers this area.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Ethics approval** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article or uploaded as supplementary information.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iD

Tafadzwa Patience Kunonga <http://orcid.org/0000-0002-6193-1365>

## REFERENCES

- 1 Matthews NH, Li WQ, Qureshi AA, *et al.* Cutaneous Melanoma: etiology and therapy. In: Ward WH, Farma JM, eds. *Epidemiology of melanoma*. Brisbane (AU): Codon Publications, 2017.
- 2 de Vries E, Sierra M, Piñeros M, *et al.* The burden of cutaneous Melanoma and status of preventive measures in central and South America. *Cancer Epidemiol* 2016;44 Suppl 1:S100–9.
- 3 Svedman FC, Pillas D, Taylor A, *et al.* Stage-specific survival and recurrence in patients with cutaneous malignant Melanoma in Europe – a systematic review of the literature. *Clin Epidemiol* 2016;8:109–22.
- 4 Balch CM, Gershenwald JE, Soong S-J, *et al.* Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol* 2009;27:6199–206.
- 5 Thomson DR, Rughani MG, Kuo R, *et al.* Sentinel node biopsy status is strongly predictive of survival in cutaneous Melanoma: extended follow-up of Oxford patients from 1998 to 2014. *J Plast Reconstr Aesthet Surg* 2017;70:1397–403.
- 6 Gershenwald JE, Ross MI. Sentinel-lymph-node biopsy for cutaneous melanoma. *N Engl J Med* 2011;364:1738–45.
- 7 Barbour A, Millward M, Morton R, *et al.* Clinical practice guidelines for the diagnosis and management of melanoma Sydney: cancer council Australia:clinical guidelines network. 2018 Available: <https://wiki.cancer.org.au/australia/Guidelines:Melanoma> [Accessed 11 Oct 2019].
- 8 American Joint Committee on Cancer. *AJCC Cancer Staging Manual* 8th ed. Chicago: American Joint Committee on Cancer/ Springer, 2017.
- 9 American Joint Committee on Cancer. *AJCC Cancer Staging Manual* 7th ed. New York, NY: Chicago: American Joint Committee on Cancer, 2010.
- 10 Moons KGM, Kengne AP, Grobbee DE, *et al.* Risk prediction models: II. external validation, model updating, and impact assessment. *Heart* 2012;98:691–8.
- 11 Usher-Smith JA, Emery J, Kassianos AP, *et al.* Risk prediction models for melanoma: a systematic review. *Cancer Epidemiol Biomarkers Prev* 2014;23:1450–63.
- 12 Vuong K, McGeechan K, Armstrong BK, *et al.* Risk prediction models for incident primary cutaneous melanoma: a systematic review. *JAMA Dermatol* 2014;150:434–44.
- 13 Kaiser I, Pfahlberg AB, Uter W, *et al.* Risk prediction models for melanoma: a systematic review on the heterogeneity in model development and validation. *Int J Environ Res Public Health* 2020;17:7919.
- 14 Mahar AL, Compton C, Halabi S, *et al.* Critical assessment of clinical prognostic tools in melanoma. *Ann Surg Oncol* 2016;23:2753–61.
- 15 Cancer Research UK. Treatment for stage 1 and 2 melanoma 2022, Available: [chrome-extension://efaidnbnmnibpcjpcglclefindmkaj/https://www.nhs.uk/ipgmedia/National/Cancer%20Research%20UK/assets/Treatingearly\(Stage1\)melanoma\(CRUK2pages\).pdf](https://www.nhs.uk/ipgmedia/National/Cancer%20Research%20UK/assets/Treatingearly(Stage1)melanoma(CRUK2pages).pdf) [Accessed 21 Jun 2022].
- 16 Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62:1006–12.
- 17 Geersing G-J, Bouwmeester W, Zuihof P, *et al.* Search filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. *PLoS One* 2012;7:e32844.
- 18 Hayden JA, Wilson MN, Riley RD, *et al.* Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor review. *Cochrane Database Syst Rev* 2019;2019:CD011284.
- 19 Whitlock EP, Lopez SA, Chang S, *et al.* AHRQ series paper 3: identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the effective health-care program. *J Clin Epidemiol* 2010;63:491–501.
- 20 Moons KGM, de Groot JAH, Bouwmeester W, *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;11:e1001744.
- 21 Wolff RF, Moons KGM, Riley RD, *et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies (prediction model risk of bias assessment tool). *Ann Intern Med* 2019;170:51–8.
- 22 de Jong Y, Ramspek CL, Zoccali C, *et al.* Appraising prediction research: a guide and meta-review on bias and applicability assessment using the prediction model risk of bias assessment tool (PROBAST). *Nephrology (Carlton)* 2021;26:939–47.
- 23 Schünemann H, Brozek J, Guyatt G, *et al.* GRADE handbook for grading quality of evidence and strength of recommendations 2013, Available: <https://gdt.gradepro.org/app/handbook/handbook.html> [Accessed 18 Dec 2020].
- 24 Iorio A, Spencer FA, Falavigna M, *et al.* Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;350:h870.
- 25 Baade PD, Royston P, Youl PH, *et al.* Prognostic survival model for people diagnosed with invasive cutaneous melanoma. *BMC Cancer* 2015;15:27.
- 26 Balch CM, Soong SJ, Gershenwald JE, *et al.* Prognostic factors analysis of 17,600 melanoma patients: validation of the American



- joint committee on cancer melanoma staging system. *J Clin Oncol* 2001;19:3622–34.
- 27 Cochran AJ, Elashoff D, Morton DL, *et al.* Individualized prognosis for melanoma patients. *Hum Pathol* 2000;31:327–31.
  - 28 El Sharouni M-A, Ahmed T, Varey AHR, *et al.* Development and validation of nomograms to predict local, regional, and distant recurrence in patients with thin (T1) melanomas. *J Clin Oncol* 2021;39:1243–52.
  - 29 Fonseca IB, Lindote MVN, Monteiro MR, *et al.* Sentinel node status is the most important prognostic information for clinical stage IIB and IIC melanoma patients. *Ann Surg Oncol* 2020;27:4133–40.
  - 30 Gimotty PA, Elder DE, Fraker DL, *et al.* Identification of high-risk patients among those diagnosed with thin cutaneous melanomas. *J Clin Oncol* 2007;25:1129–34.
  - 31 Gimotty PA, Guerry D, Ming ME, *et al.* Thin primary cutaneous malignant melanoma: a prognostic tree for 10-year metastasis is more accurate than American joint committee on cancer staging. *J Clin Oncol* 2004;22:3668–76.
  - 32 Maurichi A, Miceli R, Camerini T, *et al.* Prediction of survival in patients with thin melanoma: results from a multi-institution study. *J Clin Oncol* 2014;32:2479–85.
  - 33 Rosenbaum BE, Schafer C, Han SW, *et al.* Computer-assisted image analysis demonstrates tumor area and width as Prognostic factors in stage IB Melanoma. *Pigment Cell Melanoma Res* 2017;30.
  - 34 Soong S, Ding S, Coit D, *et al.* Predicting survival outcome of localized melanoma: an electronic prediction tool based on the AJCC melanoma database. *Ann Surg Oncol* 2010;17:2006–14.
  - 35 Tsai C-A, Chen D-T, Chen JJ, *et al.* An integrated tree-based classification approach to prognostic grouping with application to localized melanoma patients. *J Biopharm Stat* 2007;17:445–60.
  - 36 Verver D, van Klaveren D, Franke V, *et al.* Development and validation of a nomogram to predict recurrence and melanoma-specific mortality in patients with negative sentinel lymph nodes. *Br J Surg* 2019;106:217–25.
  - 37 Vollmer RT, Seigler HF. Using a continuous transformation of the Breslow thickness for prognosis in cutaneous melanoma. *Am J Clin Pathol* 2001;115:205–12.
  - 38 Xiao Y, Peng S, Hu Y, *et al.* Development and validation of prognostic nomogram in patients with nonmetastatic malignant melanoma: a SEER population-based study. *Cancer Med* 2020;9:8562–70.
  - 39 Debray TPA, Koffijberg H, Nieboer D, *et al.* Meta-Analysis and aggregation of multiple published prediction models. *Stat Med* 2014;33:2341–62.
  - 40 Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med* 2014;33:517–35.
  - 41 Steyerberg EW. *Clinical prediction models*. New York, NY: Springer, 2009.
  - 42 Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016;352:i6.
  - 43 Steyerberg EW, Harrell FE, Borsboom GJ, *et al.* Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774–81.
  - 44 Debray TPA, Vergouwe Y, Koffijberg H, *et al.* A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015;68:279–89.
  - 45 Wong T-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition* 2015;48:2839–46.
  - 46 Johansson I, Tempel D, Dwarkasing JT, *et al.* Validation of a Clinicopathological and gene expression profile model to identify patients with cutaneous Melanoma where sentinel lymph node biopsy is unnecessary. *Eur J Surg Oncol* 2022;48:320–5.
  - 47 Buja A, Bardin A, Damiani G, *et al.* Prognosis for cutaneous Melanoma by clinical and pathological profile: a population-based study. *Front Oncol* 2021;11:737399.
  - 48 Steyerberg EW, Vickers AJ, Cook NR, *et al.* Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology* 2010;21:128–38.
  - 49 Vale L, Kunonga P, Coughlan D, *et al.* Optimal surveillance strategies for patients with stage 1 cutaneous melanoma post primary tumour excision: three systematic reviews and an economic model. *Health Technol Assess* 2021;25:1–178.
  - 50 Kaiser I, Pfahlberg AB, Mathes S, *et al.* Inter-rater agreement in assessing risk of bias in melanoma prediction studies using the prediction model risk of bias assessment tool (PROBAST): results from a controlled experiment on the effect of specific rater training. *J Clin Med* 2023;12:1976.
  - 51 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015;13:1.
  - 52 Kaiser I, Diehl K, Heppt MV, *et al.* Reporting quality of studies developing and validating melanoma prediction models: an assessment based on the TRIPOD statement. *Healthcare (Basel)* 2022;10:238.
  - 53 Michaelson J. Melanoma Conditional Outcome Calculator 2011, . 2022 Available: <http://lifemath.net/cancer/melanoma/condsurv/index.php> [Accessed 13 Jun 2022].
  - 54 Michaelson J. Melanoma Outcome Calculator 2011, . 2022 Available: <http://lifemath.net/cancer/melanoma/outcome/index.php> [Accessed 13 Jun 2022].
  - 55 Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res* 2019;3:18.
  - 56 Ding L, Gosh A, Lee DJ, *et al.* Prognostic biomarkers of cutaneous melanoma. *Photodermatol Photoimmunol Photomed* 2022;38:418–34.
  - 57 Liu R, Lv Q-L, Yu J, *et al.* Correlating transcriptional networks with pathological complete response following neoadjuvant chemotherapy for breast cancer. *Breast Cancer Res Treat* 2015;151:607–18.
  - 58 Adler NR, Wolfe R, Kelly JW, *et al.* Tumour mutation status and sites of metastasis in patients with cutaneous melanoma. *Br J Cancer* 2017;117:1026–35.
  - 59 Calomarde-Rees L, García-Calatayud R, Requena Caballero C, *et al.* Risk factors for lymphatic and hematogenous dissemination in patients with stages I to II cutaneous melanoma. *JAMA Dermatol* 2019;155:679–87.
  - 60 Sadatsafavi M, Adibi A, Puhani M, *et al.* Moving beyond AUC: decision curve analysis for quantifying net benefit of risk prediction models. *Eur Respir J* 2021;58:2101186.