Cancer Epidemiology, Biomarkers & Prevention

# Risk prediction models for melanoma: A systematic review

Juliet A. Usher-Smith, Jon Emery, Angelos P. Kassianos, et al.

*Cancer Epidemiol Biomarkers Prev* Published OnlineFirst June 3, 2014.

# Risk prediction models for melanoma: A systematic review

Juliet A Usher-Smith[1], Jon Emery[1,2,3], Angelos P. Kassianos[1], Fiona M Walter[1,2,3]

[1] The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

[2] General Practice and Primary Care Academic Centre, University of Melbourne, Australia.

[3] School of Primary, Aboriginal and Rural Health Care, University of Western Australia, Australia.

**Running title:** Risk prediction models for melanoma: A systematic review

**Keywords:** Risk, prediction, melanoma, review

**Financial support:** This report is independent research arising from a Clinician Scientist award supported by the National Institute for Health Research (RG 68235) and J Usher-Smith is funded by a National Institute for Health Research Clinical Lectureship. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

**Correspondence to:** J Usher-Smith, The Primary Care Unit, University of Cambridge, Strangeways Research Laboratory, 2 Worts Causeway, Cambridge CB1 8RN. Email: jau20@medschl.cam.ac.uk Tel: 01223 748693 Fax: 01223 768412

**Conflicts of interest:** None

**Word count:** 4261

**Total number of figures and tables:** 5

**ABSTRACT**

Melanoma incidence is rising rapidly worldwide among white skinned populations. Earlier diagnosis is the principal factor that can improve prognosis. Defining high-risk populations using risk prediction models may help targeted screening and early detection approaches. In this systematic review we searched Medline, EMBASE and the Cochrane Library for primary research studies reporting or validating models to predict risk of developing cutaneous melanoma. 4141 papers were identified from the literature search and six through citation searching. 25 risk models were included. Between them, the models considered 144 possible risk factors, including 18 measures of number of naevi and 26 of sun/UV exposure. Those most frequently included in final risk models were number of naevi, presence of freckles, history of sunburn, hair colour and skin colour. Despite the different factors included and different cut-offs for sensitivity and specificity, almost all models yielded sensitivities and specificities that fit along a summary ROC with AUROC of 0.755, suggesting most models had similar discrimination. Only 2 models have been validated in separate populations and both also showed good discrimination with AUROC values of 0.79 (0.70-0.86) and 0.70 (0.64-0.77). Further research should focus on validating existing models rather than developing new ones.

## INTRODUCTION

Melanoma is one of the fastest growing cancers worldwide: age adjusted incidence rates have been increasing in most of the fair-skinned populations in recent decades; and 160000 new cases are diagnosed annually worldwide (1–5). As earlier diagnosis is the principal factor that can improve the prognosis of patients with melanoma (6), there is considerable interest in the development of screening programmes. The SCREEN project in northern Germany suggested that population screening may have a substantial impact on melanoma incidence and 5 year mortality (7, 8) leading to the implementation of a national statutory skin cancer early detection program in Germany in 2008. However, such mass screening is not currently recommended by the US Preventive Services Task Force (9) or in other countries. Modelling studies suggest that selective, targeted screening might be a more cost-effective strategy (10, 11). Such a stratified approach is currently recommended by the Royal Australian College of General Practitioners. Australian primary care physicians are advised to perform skin examinations every 3-12 months in people with multiple atypical or dysplastic naevi and a history of melanoma or a first-degree relative with melanoma (12). This approach is also being considered by the Department of Health in the United Kingdom.

The aims of such targeted screening programs are to identify people at higher risk of melanoma and to offer them preventive advice about sun protection and skin awareness and early consultation or surveillance (13–15). The identification of people at higher risk may be improved by the use of risk prediction models. Several risk models have been developed but their strengths, weaknesses and relative performance are uncertain. We report a systematic review and comparison of risk prediction models for melanoma.

## MATERIALS AND METHODS

### Search strategy

An electronic literature search of Medline, EMBASE and the Cochrane Library up to August 2013 was performed using a combination of subject headings and free text incorporating 'melanoma', 'risk/risk factor/risk assessment/chance' and 'prediction/model/score' (see Supplementary File 1 for complete search strategy). We then manually screened the reference lists of all included papers.

### Study selection

Studies were included if they fulfilled all of the following criteria: (i) are published as a primary research paper in a peer-reviewed journal; (ii) identify risk factors for developing melanoma at the level of the individual; (iii) provide a measure of relative or absolute risk using a combination of risk factors that allows identification of people at higher risk of melanoma; (iv) use a statistical method to develop the final risk model; and (iv) are applicable to the general population. As the focus of the review is to summarise the risk prediction models for incident melanoma, studies developing models for the risk of recurrence and prognostic models were excluded. Studies including only highly selected groups, for example immunosuppressed patients or those with a previous history of cancer, and conference proceedings were also excluded. The decision to only include papers which use a statistical method to develop the final risk model was made to differentiate between those studies which had set out to develop a risk model, using either a step-wise method or maximisation of sensitivity and specificity to select the variables for the final model, from the large number of variable-finding studies which provide tables with odds ratios or risk ratios adjusted simultaneously for all considered variables but do not attempt to generate or test a risk model.

One reviewer (JUS) performed the search and screened the titles and abstracts to exclude papers that were clearly not relevant. A second reviewer (FW) independently assessed a random selection of 5% of the papers excluded at that stage. For papers where a definite decision to reject could not be made based on title and abstract alone, the full-text was examined. At least two reviewers (JUS and FW/JE/AK) independently assessed all full-text papers, and those deemed not to meet inclusion criteria by both researchers were excluded. Papers for which it was unclear whether or not the inclusion criteria were met were discussed at consensus meetings including all researchers. Papers written in languages other than English were translated into English for assessment and subsequent data extraction.

**Data extraction and synthesis**

Data were extracted independently by at least two researchers (JUS and FW/JE/AK) using a standardised form to minimise bias. The form included details on: (i) the development of the model, including the study design, selection of participants, the variables considered for inclusion in the model and how they were selected; (ii) the risk model itself, including the variables included, the method of administration and whether it requires physician input or population training; (iii) the performance of the risk model in the development population, including measures of discrimination, accuracy, calibration and utility; and (iv) validation studies of the risk model and data collection tool, including the study design and performance of the risk model.

For studies which reported the step-wise performance of models, only the model with the best performance was included. For studies which included multiple different models, for example separate models for men and women or for self-assessment and physician assessment, all were included separately. One paper (16) reported models for two different

age groups in additional to the cohort as a whole. In this case only the model for the entire cohort was included.

During the data extraction, risk factors were grouped into the following categories: personal characteristics; genetic factors; female hormonal factors; access to specialist skin care; personal medical history; family history; hair colour; eye colour; skin type (Fitzpatrick); skin colour, skin response to sun; history of sunburn; use of sun protection; number of naevi; number of atypical or dysplastic naevi; freckles; congenital naevi; other skin findings; new or changing naevi; sun/UV exposure (including sun bed use); and UV skin damage. Separate categories were included for skin colour, skin response to the sun, and skin type (Fitzpatrick), which includes both skin colour and skin response to the sun (17). If papers used the term 'skin type' but then defined that by the skin response to the sun this was extracted under skin response to the sun.

Information concerning whether the risk models required physician input or could be performed without involvement of a healthcare professional was also extracted. Risk models were classified as requiring physician input if they included any of the following factors: dysplastic or atypical naevi; actinic lentigines; total body naevus count; genetic analysis requiring samples; or specialised equipment such as dermoscopy or colorimetry. Naevus density, as in Marrett et al 1992 (18), was not considered to require physician input as participants were provided with images representing a range of naevus density and counting of individual naevi was not required.

Reported measures of discrimination, accuracy, calibration and utility were used to compare the performance of risk models. The sensitivity and specificity of different models was also compared graphically by plotting a summary ROC curve using the Moses-Littenberg method (19, 20) in RevMan version 5.2 and the summary AUROC calculated in

Meta-DiSc version 1.4 using Moses' constant for linear models to fit the summary ROC curve.

## RESULTS

After duplicates were removed, the search identified 4141 papers. 4080 of these were excluded at title and abstract level. A further 42 were excluded after full-text assessment by at least two authors (JUS and FW/JE/AK). There was complete agreement amongst researchers throughout the screening process and the most common reasons for exclusion were that the papers did not use a statistical method to develop the final risk model, were conference abstracts or not primary research. Two well cited models, excluded because they were not developed using a statistical method, are those by Mar (21) and Glanz (22). Mar et al 2011 selected risk factors and estimated relative risks for risk factor combinations from existing large meta-analyses (23–25) and data from the Victorian Cancer Registry. Glanz developed the BRAT (Brief skin cancer Risk Assessment Tool) through critically reviewing published literature on risk factors and their self-assessment and then piloted the questionnaire on a convenience sample of people at varying levels of risk to estimate the range of scores and test-retest reliability of the tool. Neither tested the performance of the models in any populations with melanoma.

Six further papers were identified through citation searching giving 25 papers for inclusion. Of these, four provided validation of other models and four included more than one risk model. This review, therefore, describes 25 risk prediction models (Figure 1).

A summary of these 25 models, along with measures of performance in the development population and notable strengths and weaknesses, are given in Table 1. Fifteen require physician input whilst 10 can be performed by self-assessment. Discriminatory performance was provided for fourteen. Most had values for the area under the receiver

operating curve (AUROC) between 0.7-0.8 with little difference between those suitable for self-assessment and those requiring a health care professional. Poorer discrimination was seen in those models including only skin colour and skin type (0.54) (26), age, sex, cutaneous melanin and genotyping (0.65) (27) and in the only model developed in a cohort study to provide a measure of performance (0.62) (28). The highest discrimination was for a model including a suspicious melanoncytic lesion on dermoscopy (0.86) (29) with a second model developed from a small case control study in Brazil where there were more cases than controls also reporting high discrimination (0.85) (30).

A measure of accuracy was provided in ten studies. The sensitivity and specificity varied between them but a summary ROC curve (Figure 2) shows that they all lie very close to the curve. This shows that despite all the heterogeneity in model development and risk factors, there is very little heterogeneity in the predictive ability of the models with the variation in sensitivity and specificity likely a reflection of the cut offs chosen in different studies. The AUROC of this summary curve is 0.755.

Only three models had reported measures of calibration (28, 29, 31). All three showed good calibration but all had been tested in the development population where calibration would be expected to be high.

Further details of the development of each model are given in Supplementary Tables S1 and S2 for case-control and cohort studies respectively. Twenty-one were case-control studies and 4 cohort studies. Overall the reporting of the studies was variable. Of the 21 case-control studies, the method of selecting the variables for consideration was given in only 11, of which for 8 the method was a literature review, and the predictor variables and outcomes were evaluated in a blinded fashion in only 4. Cases were selected from either cancer registries or dermatology clinics and all required histological confirmation of diagnosis. Controls were selected from hospital clinics in 7 studies, the general population in

5, dermatology clinics in 4 and primary care in 3. Most controls were matched by age and gender with a mean age of 43-57 years. Of the 4 cohort studies, only Neilsen et al 2011 (16) provided any detail of the method of selecting the variables, none were evaluated in a blinded fashion and Goldberg 2007 (32) did not require a histological diagnosis. Neilsen et al 2011 (16) included only female participants with Cho et al 2005 (28) also heavily female dominated.

Table 2 shows additional details of those models in which either the model itself or the method of data collection used for the model has been validated or in which efficiency has been estimated. Only one model, Fortes et al 2010 (33), has been validated in an external population and one, Williams et al 2011 (34), in a separate sub-group of the original study population. English et al (35) also divided their initial study population into two but they used the second sub-population to further refine the model developed in the first sub-population rather than validate the performance of the model in a separate population.

Between them, the 25 risk prediction models considered 144 different possible risk factors (Supplementary Table S3). These included 18 different measures of number of naevi, 26 of sun / UV exposure and 14 of history of sunburn. There were also multiple definitions of dysplastic naevi with each research group using a different definition. Categorising the different risk factors, as shown in Supplementary Table S3, allowed comparison of those considered and included in each of the risk models (Table 3). This shows that the risk factors most frequently included in the models are (in order of frequency) number of naevi, freckles, hair colour, skin colour, history of sunburn and sun/UV exposure. The risk factors most likely to remain in the final model after consideration are age, number of naevi, skin type, skin colour, personal history of skin cancer and freckles. Ethnicity, other personal characteristics such as socio-demographic measures, female hormonal factors, use of sun protection and congenital naevi were not included in any of the final models and a family

history of skin cancer and eye colour were included in the final model in less than one in five times they were considered.

## DISCUSSION

### Principal findings

This is the first systematic review of risk prediction models for melanoma. It shows that multiple risk models exist and that they have the potential to identify individuals at higher risk of melanoma. Comparisons between the different models are difficult due to the lack of validation studies and heterogeneity in choice and definition of variables. Despite this, however, we show that most include well established risk factors and the AUROC of a summary ROC curve is comparable with those for other cancers, such as breast cancer (0.716-0.762) (36) and colon cancer (0.61-0.74) (37). There was also little difference in model performance between those scores suitable for self-assessment and those requiring a health care professional, suggesting potential for use at a population level to identify people at higher risk of melanoma.

### Strengths and limitations

The main strengths of this review are the use of broad inclusion criteria and the systematic search of multiple databases not limited by language. This approach enabled us to identify published risk models even when developing the risk model had not been the primary aim of the study, and in doing so reduces the risk of selection bias. Whilst we cannot exclude publication bias we also expect this to be minimal because of the exploratory nature of many of the studies and the absence of performance data.

As with most systematic reviews, the main limitation is the quality of the published data. Notably, in this review it was difficult to perform direct comparisons of the risk models

due to the lack of validation studies for most of the risk models. The majority of studies also gave no indication of how the authors decided which risk factors to consider for inclusion in the model and 144 different risk factors were considered with varying definitions. Additionally, many of the risk factors are subjective in nature and subject to recall bias, which is likely to overestimate the performance of those models developed from case control studies, and only 4 included blinding of the investigator to melanoma status. By presenting all the risk models together for the first time, however, we are able to demonstrate this heterogeneity whilst making comparisons where possible.

**Evaluation of the risk models**

The 25 risk models differ in the risk factors included, the method of administration and their performance. Most contain established risk factors for melanoma, however, there was considerable variation amongst the definitions and measures used. In some cases, notably history of sunburn and sun / UV exposure, this likely reflects the difficulty measuring exposure to the risk factor, both due to its subjective nature and the need to recall events in the past. This is in contrast to more objective and consistent measures, such as eye colour, skin type or hair colour for which many fewer variations were seen. In other cases, particularly number of naevi and atypical and dysplastic naevi, the range of definitions probably reflects on-going uncertainty within the literature and the controversy around a non-histological diagnosis of an atypical or dysplastic naevus (38). In all cases, however, it demonstrates the large number of variables in use within the field. Whilst it is unlikely that a single measure of each risk factor will be appropriate for all situations, increased consistency would allow more meaningful comparisons in future research.

With such a large number of risk factors considered it is perhaps not surprising that the models differ widely in the risk factors included. Most include a measure of number of

naevi and skin type or colour and either include or adjust for age and gender but beyond that it is difficult to make generalisations.

Performance measures were only available for 16 models in the development population and two in external populations. Despite the variations already described, however, the accuracy, measured by the sensitivity and specificity, is consistent across them. By virtue of the cut-offs set by the authors, some have higher specificity and lower sensitivity (18, 34, 35, 39) whilst others have higher sensitivity and lower specificity (26, 31, 33). The summary ROC curve, however, shows that, despite including a range of different variables, there is very little heterogeneity in the predictive ability of the models with the variations in sensitivity and specificity reflecting different cut-offs. One reason for this may be that there is a group of core risk factors responsible for most of the increased risk. Due to the range of factors included in the different models, however, it is not possible to identify those from the available studies.

The discrimination of the models, as measured by the area under receiver operating curve, compares favourably with risk models used for other cancers, including breast cancer with AUROCs of 0.716-0.762 (36) and colon cancer with AUROCs of 0.61-0.74 (37). Care must be taken when making such comparisons, however, as many of these have been developed and validated in large cohort studies whilst the majority of melanoma risk models have been developed from case-control studies with up to 60% prevalence of melanoma which will inflate their performance through spectrum bias.

**Evaluation of individual risk factors**

Whilst evaluation of individual melanoma risk factors was not the primary aim of this study, by including only studies that used a statistical method to develop a risk model and extracting the number of times a risk factor was included in the final model when it was

considered, the results of our analysis confirm the importance of several established risk factors for melanoma (23–25). These include age, number of naevi, skin type and colour, personal history of melanoma or non-melanocytic skin cancer, freckles, dysplastic naevi and hair colour. Sun exposure, history of sunburn and skin response to the sun were also included in many of the final models but only half the times they were considered (53, 50 and 47% respectively), perhaps reflecting their subjective nature and risk of recall bias. Eye colour was also only included in 4 of the 13 models in which it was considered and this is likely to be due to known correlation between hair colour, eye colour, freckles and skin colour (23).

An unexpected finding was the absence of family history in many of the models. It was considered in 18 of the models but only remained in the final score in 6. This differs from earlier studies in which approximately 10% of cases of melanoma have reported heredity (23, 40). It may be that other phenotypic markers which remain in the risk model are strongly correlated with family history or may simply reflect the very low incidence of true familial melanoma in the melanoma population. Some other risk factors, including for example genetic factors, ethnicity and female hormone factors, were also not considered by very many of the models and so their potential importance may be underestimated.

**Implications for clinicians and policy makers**

This review shows that multiple risk models for prediction of the development of melanoma exist and that they have the potential to identify individuals at higher risk of melanoma. Clinicians will be interested to see the range and relative performance of different risk models. However, all the risk scores were developed to predict risk of future disease rather than undiagnosed prevalent disease. Consequently, the results of this review will be of particular relevance to policy makers interested in the potential for using risk scores among asymptomatic people to identify a subset of the population for whom targeted

screening, surveillance or educational programmes could be offered to reduce the morbidity and mortality from melanoma.

As English and Armstrong (35) point out, if a screening programme is to be directed towards a high risk group and is to have an impact on the disease as a whole, three criteria must be satisfied in addition to those for all screening programmes (41): People at high risk of the disease must be readily identifiable; those identified as being at high risk must form a large proportion of all patients who develop the disease; and this proportion must be substantially larger than the proportion of the whole population that constitutes the group at high risk. When assessed against these three criteria this review confirms that risk models exist which could be used to identify a group at higher risk of melanoma. Firstly, a number of risk models exist for which patient self-assessment is feasible and so they could be undertaken in clinical waiting rooms or via on-line platforms (16, 18, 28, 30–32, 34, 35, 39). Secondly, those models that provide values for sensitivity and specificity suggest that screening could identify a high risk group containing between 25-89% of people expected to develop melanoma and, thirdly, that this high risk group would comprise between 10-55% of the population. These ranges are wide due to the variation in cut-offs selected in each study and reflects the trade-off between sensitivity and specificity. For example, from the summary ROC, choosing a risk score with a specificity of 50% when 50% of the population would be classified as higher risk, sensitivity is around 80% so 80% of melanomas would be detected in that higher risk group. Choosing instead a score with a specificity of 80% when 20% would identified as high risk, the sensitivity falls to around 50% so only 50% of cases would be detected.

Some, including Fortes (33), believe that as melanoma can be a fatal disease but referral to a dermatologist and excision or biopsy is relatively benign, it is better to give priority to sensitivity over specificity as the inclusion of false-positive cases may be less

detrimental than false-negatives. However, the utility of a risk score that identifies 50% of the population as higher risk is limited and any screening of asymptomatic people has considerable implications in terms of health care costs and both physical and psychological consequences. Several previous studies have estimated the cost-effectiveness of various melanoma screening strategies. One-off screening of a white population of all ages at average risk by a dermatologist has been shown to cost $172,276 per year-of-life-saved (YLS) (42) but this cost falls dramatically when screening is targeted to higher risk populations, defined variously by age, family history or phenotypic characteristics (10, 11, 43). Whilst a full economic analysis is beyond the scope of this review, the risk scores described are able to identify higher risk groups with greater discriminatory ability and accuracy than age, family history or phenotypic characteristics alone, and so any screening programme based on one of the risk models is likely to be even more cost-effective.

**Implications for future research**

The finding that many of the models have similar performance characteristics despite the wide range of different variables included suggests that developing further models based on current known risk factors is unlikely to benefit the field. As advances are made into genes that play a role in the susceptibility of melanoma (44, 45), development of new risk models incorporating genetic information may improve the discriminatory ability. Until then, further research should focus on validating existing models in different populations and assessing the costs, feasibility, acceptability and adverse consequences of applying these models.

**ACKNOWLEDGEMENTS**

We thank Isla Kuhn, Reader Services Librarian, University of Cambridge Medical Library, for her help developing the search strategy and Professor Simon Griffin and the MelaTools programme's expert panel (Dr Katharine Acland, Dr Nigel Burrows, Dr Pippa Corrie and Mr Per Hall) for their helpful comments on the manuscript.

**REFERENCES**

1. Jemal, A, Devesa, SS, Hartge, P, Tucker, MA. Recent trends in cutaneous melanoma incidence among whites in the United States. J. Natl. Cancer Inst. 2001; 93: 678–83.

2. Vries, E de, Coebergh, J-WW. Melanoma incidence has risen in Europe. BMJ 2005; 331: 698.

3. Coory, M, Baade, P, Aitken, J, Smithers, M, McLeod, GRC, Ring, I. Trends for in situ and invasive melanoma in Queensland, Australia, 1982-2002. Cancer Causes Control 2006; 17: 21–7.

4. Downing, A, Yu, XQ, Newton-Bishop, J, Forman, D. Trends in prognostic factors and survival from cutaneous melanoma in Yorkshire, UK and New South Wales, Australia between 1993 and 2003. Int. J. Cancer 2008; 123: 861–6.

5. Linos, E, Swetter, SM, Cockburn, MG, Colditz, GA, Clarke, CA. Increasing burden of melanoma in the United States. J. Invest. Dermatol. 2009; 129: 1666–74.

6. 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; 27: 6199–206.

7. Waldmann, A, Nolte, S, Weinstock, MA, Breitbart, EW, Eisemann, N, Geller, AC, et al. Skin cancer screening participation and impact on melanoma incidence in Germany--an observational study on incidence trends in regions with and without population-based screening. Br. J. Cancer 2012; 106: 970–4.

8. 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; 66: 201–11.

9. Wolff, T, Tai, E, Miller, T. Screening for skin cancer: an update of the evidence for the U.S. Preventive Services Task Force. Ann. Intern. Med. 2009; 150: 194–8.

10. Freedberg, KA, Geller, AC, Miller, DR, Lew, RA, Koh, HK. Screening for malignant melanoma: A cost-effectiveness analysis. J. Am. Acad. Dermatol. 1999; 41: 738–45.

11. Losina, E, Walensky, RP, Geller, A, Beddingfield, FC, Wolf, LL, Gilchrest, BA, et al. Visual screening for malignant melanoma: a cost-effectiveness analysis. Arch. Dermatol. 2007; 143: 21–8.

12. Royal Australian College of General Practitioners. Guidelines for preventive activities in general practice 8th edition.

13. Masri, GD, Clark, WH, Guerry, D, Halpern, A, Thompson, CJ, Elder, DE. Screening and surveillance of patients at high risk for malignant melanoma result in detection of earlier disease. J. Am. Acad. Dermatol. 1990; 22: 1042–8.

14. Wolfe, JT. The role of screening in the management of skin cancer. Curr. Opin. Oncol. 1999; 11: 123–8.

15. Eiser, JR, Pendry, L, Greaves, CJ, Melia, J, Harland, C, Moss, S, et al. Is targeted early detection for melanoma feasible? Self assessments of risk and attitudes to screening. J. Med. Screen. 2000; 7: 199–202.

16. Nielsen, K, Masback, A, Olsson, H, Ingvar, C. A prospective, population-based study of 40 000 women regarding host factors, UV exposure and sunbed use in relation to risk and anatomic site of cutaneous melanoma. Pigment Cell Melanoma Res. 2011; 24: 1071.

17. Fitzpatrick, TB. The validity and practicality of sun-reactive skin types I through VI. Arch. Dermatol. 1988; 124: 869–71.

18. Marrett, LD, King, WD, Walter, SD, From, L. Use of host factors to identify people at high risk for cutaneous malignant melanoma . CMAJ 1992; 147: 445–453.

19. Littenberg, B, Moses, LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med. Decis. Making 13: 313–21.

20. Moses, LE, Shapiro, D, Littenberg, B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat. Med. 1993; 12: 1293–316.

21. Mar, V, Wolfe, R, Kelly, JW. Predicting melanoma risk for the Australian population. Australas. J. Dermatol. 2011; 52: 109–116.

22. Glanz, K, Schoenfeld, E, Weinstock, MA, Layi, G, Kidd, J, Shigaki, DM. Development and reliability of a brief skin cancer risk assessment tool. Cancer Detect. Prev. 2003; 27: 311–315.

23. Gandini, S, Sera, F, Cattaruzza, MS, Pasquini, P, Zanetti, R, Masini, C, et al. Meta-analysis of risk factors for cutaneous melanoma: III. Family history, actinic damage and phenotypic factors. Eur. J. Cancer 2005; 41: 2040–59.

24. Gandini, S, Sera, F, Cattaruzza, MS, Pasquini, P, Picconi, O, Boyle, P, et al. Meta-analysis of risk factors for cutaneous melanoma: II. Sun exposure. Eur. J. Cancer 2005; 41: 45–60.

25. Gandini, S, Sera, F, Cattaruzza, MS, Pasquini, P, Abeni, D, Boyle, P, et al. Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi. Eur. J. Cancer 2005; 41: 28–44.

26. Barbini, P, Cevenini, G, Rubegni, P, Massai, MR, Flori, ML, Carli, P, et al. Instrumental measurement of skin colour and skin type as risk factors for melanoma: a statistical classification procedure. Melanoma Res. 1998; 8: 439–447.

27. Dwyer, T, Stankovich, JM, Blizzard, L, FitzGerald, LM, Dickinson, JL, Reilly, A, et al. Does the addition of information on genotype improve prediction of the risk of melanoma and nonmelanoma skin cancer beyond that obtained from skin phenotype? Am. J. Epidemiol. 2004; 159: 826–833.

28. Cho, E, Rosner, BA, Feskanich, D, Colditz, GA. Risk factors and individual probabilities of melanoma for whites. J. Clin. Oncol. 2005; 23: 2669–2675.

29. Guther, S, Ramrath, K, Dyall-Smith, D, Landthaler, M, Stolz, W. Development of a targeted risk-group model for skin cancer screening based on more than 100,000 total skin examinations. J. Eur. Acad. Dermatology Venereol. 2011; 26: 86–94.

30. Bakos, L, Mastroeni, S, Bonamigo, RR, Melchi, F, Pasquini, P, Fortes, C. A melanoma risk score in a Brazilian population. An. Bras. Dermatol. 2013; 88: 226–232.

31. Quereux, G, Moyse, D, Lequeux, Y, Jumbou, O, Brocard, A, Antonioli, D, et al. Development of an individual score for melanoma risk. Eur. J. Cancer Prev. 2011; 20: 217–224.

32. Goldberg, MS, Doucette, JT, Lim, HW, Spencer, J, Carucci, JA, Rigel, DS. Risk factors for presumptive melanoma in skin cancer screening: American Academy of Dermatology

National Melanoma/Skin Cancer Screening Program experience 2001-2005. J. Am. Acad. Dermatol. 2007; 57: 60–66.

33. Fortes, C, Mastroeni, S, Bakos, L, Antonelli, G, Alessandroni, L, Pilla, MA, et al. Identifying individuals at high risk of melanoma: A simple tool. Pigment Cell Melanoma Res. 2010; 19: 393–400.

34. Williams, LH, Shors, AR, Barlow, WE, Solomon, C, White, E. Identifying Persons at Highest Risk of Melanoma Using Self-Assessed Risk Factors. J. Clin. Exp. Dermatol. Res. 2011; 2: pii:1000129.

35. English, DR, Armstrong, BK. Identifying people at high risk of cutaneous malignant melanoma: Results from a case-control study in Western Australia. Br. Med. J. (Clin. Res. Ed). 1988; 296: 1285–1288.

36. Amir, E, Freedman, OC, Seruga, B, Evans, DG. Assessing women at high risk of breast cancer: a review of risk assessment models. J. Natl. Cancer Inst. 2010; 102: 680–91.

37. Win, AK, Macinnis, RJ, Hopper, JL, Jenkins, MA. Risk prediction models for colorectal cancer: a review. Cancer Epidemiol. Biomarkers Prev. 2012; 21: 398–410.

38. Farber, MJ, Heilman, ER, Friedman, RJ. Dysplastic nevi. Dermatol. Clin. 2012; 30: 389–404.

39. Harbauer, A, Binder, M, Pehamberger, H, Wolff, K, Kittler, H. Validity of an unsupervised self-administered questionnaire for self-assessment of melanoma risk. Melanoma Res. 2003; 13: 537–542.

40. Ford, D, Bliss, JM, Swerdlow, AJ, Armstrong, BK, Franceschi, S, Green, A, et al. Risk of cutaneous melanoma associated with a family history of the disease. The International Melanoma Analysis Group (IMAGE). Int. J. Cancer 1995; 62: 377–81.

41. Wilson, JM, Jungner, YG. Principles and practice of mass screening for disease. Bol. Oficina Sanit. Panam. 1968; 65: 281–393.

42. Beddingfield, F. Melanoma: a decision analysis to estimate the effectiveness and cost-effectiveness of screening and an analysis of the relevant epidemiology of the disease [dissertation]. Santa Monica (CA): RAND Graduate School; 2002

43. Girgis, A, Clarke, P, Burton, RC, Sanson-Fisher, RW. Screening for melanoma by primary health care physicians: a cost-effectiveness analysis. J. Med. Screen. 1996; 3: 47–53.

44. Newton Bishop, JA, Bishop, DT. The genetics of susceptibility to cutaneous melanoma. Drugs of today 2005; 41: 193–203.

45. Marzuka-Alcalá, A, Gabree, MJ, Tsao, H. Melanoma susceptibility genes and risk assessment. Methods Mol. Biol. 2014; 1102: 381–93.

46. Augustsson, A. Melanocytic naevi, melanoma and sun exposure. Acta Derm. Venereol. Suppl. (Stockh). 1991; 166: 1–34.

47. Fears, TR, Guerry, D, Pfeiffer, RM, Sagebiel, RW, Elder, DE, Halpern, A, et al. Identifying individuals at high risk of melanoma: a practical predictor of absolute risk. J. Clin. Oncol. 2006; 24: 3590–3596.

48. Garbe, C, Krüger, S, Stadler, R, Guggenmoos-Holzmann, I, Orfanos, CE. Markers and relative risk in a German population for developing malignant melanoma. Int. J. Dermatol. 1989; 28: 517–23.

49. Garbe, C, Büttner, P, Weiss, J, Soyer, HP, Stocker, U, Krüger, S, et al. Risk factors for developing cutaneous melanoma and criteria for identifying persons at risk: multicenter case-control study of the Central Malignant Melanoma Registry of the German Dermatological Society. J. Invest. Dermatol. 1994; 102: 695–9.

50. Landi, MT, Baccarelli, A, Calista, D, Pesatori, A, Fears, T, Tucker, M, et al. Combined risk factors for melanoma in a Mediterranean population. Br. J. Cancer 2001; 85: 1304–10.

51. Mackie, RM, Freudenberger, T, Aitchison, TC. Personal risk-factor chart for cutaneous melanoma. Lancet 1989; 2: 487–490.

52. Weiss, J, Garbe, C, Bertz, J, Biltz, H, Burg, G, Hennes, B, et al. Risk factors for the development of malignant melanoma in West Germany. Results of a multicenter-case control study. Hautarzt. 1990; 41: 309–13.

53. Zaridze, DG, Mukeriia, AF, Basieva, TK, Shlenskaia, IN, Bukin, IV. The role of endogenous and exogenous factors in the etiology of skin melanoma. Vopr. Onkol. 1992; 38: 141–147.

54. Jackson, A, Wilkinson, C, Ranger, M, Pill, R, August, P. Can primary prevention or selective screening for melanoma be more precisely targeted through general practice? A prospective study to validate a self administered risk score. BMJ 1998; 316: 34–38.

55. Westerdahl, J, Anderson, H, Olsson, H, Ingvar, C. Reproducibility of a self-administered questionnaire for assessment of melanoma risk. Int. J. Epidemiol. 1996; 25: 245–51.

56. Quereux, G, Nguyen, J-M, Volteau, C, Lequeux, Y, Dreno, B. Creation and test of a questionnaire for self-assessment of melanoma risk factors. Eur. J. Cancer Prev. 2010; 19: 48–54.

57. Quereux, G, Nguyen, J-M, Cary, M, Jumbou, O, Lequeux, Y, Dreno, B. Validation of the Self-Assessment of Melanoma Risk Score for a melanoma-targeted screening. Eur. J. Cancer Prev. 2012; 21: 588–595.

## FIGURE LEGENDS

**Figure 1**. PRISMA Flow Diagram

**Figure 2**. Summary ROC curve for the ten models that provide values for sensitivity and specificity. Each data point represents the sensitivity and specificity for a single threshold of a risk model with horizontal and vertical bars indicating 95% confidence intervals. The AUROC of this summary curve is 0.755.

Table 1. Summary of risk prediction models

| Author | Components of score | | Model performance in development population | | | General comments | |
|---|---|---|---|---|---|---|---|
| | Factors included in score | Physician input? | Discrimination | Calibration | Accuracy | Strengths | Limitations |
| Augustsson 1991 (46) | Skin type, hair colour, eye colour, total body naevus ≥ 2mm count, number of dysplastic naevi | Yes | | | | 1) Reproducible as relies on observation rather than recall | 1) Developed from survivors of melanoma so may be biased towards less poor outcomes and lower stages<br>2) Only applicable to people 30-50 years of age |
| Bakos 2013 (30) | Hair colour; presence of freckles; sunburns in all life; skin colour; eye colour | No | AUROC 0.85 (0.77-0.91) | | | 1) Good discrimination | 1) Developed from population with limited range of skin phototypes |
| Barbini 1998 (26) | Skin colour using colorimeter and skin type | Yes | AUROC 0.54 | | Sens 86; Spec 45<br>PPV 0.3; NPV 0.92 | | 1) Very complicated to calculate<br>2) Poor discrimination with two thirds of subjects misclassified as high risk |
| Cho 2005 (28) | Gender; age; family history of melanoma; history of severe sunburn; number of naevi > 3mm on arms or lower legs; hair colour | No | AUROC 0.62 (0.58-0.65) | $Chi^2$ goodness-of-fit 9.28; p = 0.41 | | 1) Based on large cohort study with 16 years follow-up | 1) Based on predominantly female white health professionals only;<br>2) Would require a computer or expert to calculate risk using regression coefficients |
| Dwyer 2004 (27) | Age; sex; cutaneous melanin; MC1R genotype | Yes | AUROC 0.65 | | | | 1)Requires DNA sample;<br>2) Only obtained genetic information from 67% of participants |
| English 1988 (35) | Number of raised naevi on the arms; age on arrival in Australia; history of non-melanocytic skin cancer; mean time spent outdoors in summer from the age of 10 to 24; family history of melanoma | No | | | Sens 54; Spec 84 | 1) High specificity<br>2) Initially developed in 400 case control pairs then refined in separate 111 pairs<br>3) Developed from large study with population based controls | 1) One variable not transferable outside Australia and number of hours spent outdoors difficult to estimate |
| Fears 2006 (men) (47) | Skin colour; number of moles < 5mm; freckling; number of moles ≥ 5mm; severe solar skin damage | Yes | AUROC 0.7-0.8 (a) | | | 1) Simple and quick with only 2 questions and examination of back | 1) Not applicable to people with prior melanoma or non-melanoma skin cancer or 1st degree relative with melanoma |
| Fears 2006 (women) (47) | Skin colour; number of moles < 5mm; freckling; tanning ability; number of moles ≥ 5mm; severe solar skin damage | Yes | AUROC 0.7-0.8 (a) | | | 1) Simple and quick with only 2 questions and examination of back | 1) Not applicable to people with prior melanoma or non-melanoma skin cancer or 1st degree relative with melanoma |
| Fortes 2010 (33) | Hair colour; skin type; presence of freckles; number of common naevi on the whole body; sunburn as a child | Yes | AUROC 0.79 (0.75-0.82) | | Risk cut off ≥ 3: Sens 88.6; Spec 51.4** | 1) Externally validated<br>2) Good discrimination and high sensitivity | 1) Developed from study with hospital based controls<br>2) Potential for recall bias with sunburn as a child |

| Author | Components of score | | Model performance in development population | | | General comments | |
|---|---|---|---|---|---|---|---|
| | Factors included in score | Physician input? | Discrimination | Calibration | Accuracy | Strengths | Weaknesses |
| Garbe 1989 (48) | Total number of naevi ≥ 2mm, total number of atypical naevi, actinic lentigines, occupational sun exposure and skin response to sun | Yes | | | | | 1) Requires physician whole body examination by dermatologist so not feasible for primary care |
| Garbe 1994 (49) | Number of naevi ≥ 2mm; presence of actinic lentigines; number of atypical melanocytic naevi; skin type; growth of any existing melanocytic naevi, hair colour | Yes | | | | 1) Reproducible as relies on observation rather than recall | 1) Hospital based controls from dermatology department;<br>2) Requires physician whole body examination by dermatologist so not feasible for primary care |
| Goldberg 2007 (32) | History of previous melanoma, age over 50, does not see regular dermatologist, changing mole, gender | No | | | | 1) Reproducible as relies on observation rather than recall | 1) Based on study with no follow up of patients and no histological diagnosis;<br>2) Risk of bias as developed from self-selected population<br>2) Questionable relevance of absent dermatologist outside USA |
| Guther 2011 (29) | Age; hair colour; past history of melanoma; suspicious melanocytic lesion on dermatoscopy | Yes | AUROC 0.86 | $Chi^2$ Likelihood ratio $p < 0.0001$ | Sens 92.3 | 1) High discrimination and calibration;<br>2) Reproducible as relies on observation rather than recall | 1) Requires dermatoscopic examination<br>2) Risk of bias as developed from self-selected population attending dermatologist |
| Harbauer 2003 (physician assessment) (39) | Skin type; UV damage to skin; number of naevi | Yes | AUROC 0.77 (0.73-0.83) | | Sens 42 (95% CI 33-52); Spec 90 | 1) Simple<br>2) Good discrimination | 1) Would require a computer to implement risk model;<br>2) Unclear how UV score was calculated<br>3) Potential for bias as developed from population with controls from private GP or dermatology |
| Harbauer 2003 (self-assessment) (39) | Skin type; UV damage to skin; number of naevi | No | AUROC 0.73 (0.6-0.77) | | Sens 39 (95% CI 31-48); Spec 90 | 1) Simple<br>2) Good discrimination | 1) Would require a computer to implement risk model<br>2) Potential for bias as developed from population with controls from private GP or dermatology |
| Landi 2001 (50) | Presence of dysplastic naevi; skin colour; propensity to tan; eye colour | Yes | | | | 1) Reproducible as relies on observation rather than recall | 1) Developed from population with most controls friends or family members of cases so potential for bias |
| MacKie 1989 (51) | Gender, total number of naevi ≥ 2mm diameter; freckling tendency; number of clinically atypical naevi ; number of episodes of severe sunburn at any time in life | Yes | | | | 1) Relatively simple to use flow chart | 1) Potential for recall bias with number of episodes of severe sunburn at any time in life |

| Author | Components of score | | Model performance in development population | | | General comments | |
|---|---|---|---|---|---|---|---|
| | Factors included in score | Physician input? | Discrimination | Calibration | Accuracy | Strengths | Weaknesses |
| Marrett 1992 (18) | Hair colour; skin reaction to repeated sun exposure; freckle density; naevus density | No | | | Sens 40; Spec 89 | 1) Limited opportunity for recall bias | 1) Not applicable to patients with previous melanoma |
| Neilsen 2011 (16) | Family history, number of naevi ≥ 3mm on left arm, hair colour, time spent on sunbathing vacations | No | | | | 1) Developed from population based cohort study | 1) Only applies to women<br>2) Based on small number of cases as relatively short period of follow up |
| Quereux 2011 (1) (31) | Sunburn in childhood; family history of melanoma; number of naevi on arms; density of freckles; skin type; total sun exposure | No | AUROC 0.70 | | Risk cut off 24: Sens 60.2 ± 2.8; Spec 71.1 ± 1.2 | 1) Good discrimination | 1) Total sun exposure difficult to calculate<br>2) Potential for recall bias with sunburn in childhood |
| Quereux 2011 (2) (31) | Sex; age; skin type; freckles; number of naevi on arms; severe blistering sunburn as a child; life in a country at low altitude; melanoma in a first degree relative. | No | AUROC 0.73 | Hosmer-Lemeshow statistic p = 0.43 (d) | Risk cut off 13: Sens 64.9 ± 3.4; Spec 68.4 ± 1.3 | 1) Good discrimination | 1) Potential for recall bias with sunburn as a child |
| Quereux 2011 (3) (31) | SAMScore (c): phototype I or II; freckling tendency; > 20 naevi on both arms; severe sunburn during childhood or teenage years; life in a country at low altitude; a history of previous melanoma; history of melanoma in a first-degree relative | No | AUROC 0.71 | | Sens 63.2 ± 3.6; Spec 68.8 ± 1.2 | 1) Good discrimination | 1) Combinatorial analysis quite complicated<br>2) Potential for recall bias with sunburn as a child |
| Weiss 1990 (52) | Total number of naevi > 2mm over whole body, hair colour, occupational sun exposure and skin response to sun | Yes | | | | | 1) Unclear description of variables included |
| Williams 2011 (34) | Age; sex; number of severe sunburns aged 2-18; hair colour age 15; density of freckles on arms before aged 20; number of raised moles on both arms; prior non-melanoma skin cancer | No | AUROC 0.77 (0.73-0.81)<br>AUROC 0.70 (0.64-0.77) (b) | | Cut off 25: Sens 61; Spec 80 (b)<br>Cut off 28: Sens 50; Spec 85 (b)<br>Cut off 30: Sens 42; Spec 90 (b)<br>Cut off 34: Sens 29; Spec 95 (b) | 1) Good discrimination<br>2) Validated on separate group<br>3) Only counting raised moles distinguishes from freckles | 1) Not yet validated for self-completion<br>2) Only applicable ages 35-74<br>3) Potential recall bias for number of freckles before age 20 |
| Zaridze 1992 (53) | Presence of freckles on arms; number of raised moles on arms and moles on body > 6mm; skin colour; eye colour; frequency of sunbathing during lifetime | No | | | | 1) Only counting raised moles distinguishes from freckles | 1) Poor description of variables included<br>2) Potential for recall bias with frequency of sunbathing during lifetime |

(a) Uses US Surveillance, Epidemiology and End Results Programme (SEER) and hypothetical cohort rather than testing on this case-control study population

(b) From validation study in different population to development of the model

(c) According to the SAMScore, a patient is considered at risk of melanoma if at least one of these 3 criteria is verified: First criterion: Presence of at least 3 risk factors among the 7 following risk factors: phototype I or II, freckling tendency, number of melanocytic naevi >20 on both arms, severe sunburn during childhood or teenage years, life in a country at low latitude, a history of previous melanoma, a history of melanoma in a first-degree relative Second criterion: A subject under 60 years of age and a number of melanocytic naevi >20 on both arms Third criterion: A subject of 60 years old or over and a freckling tendency

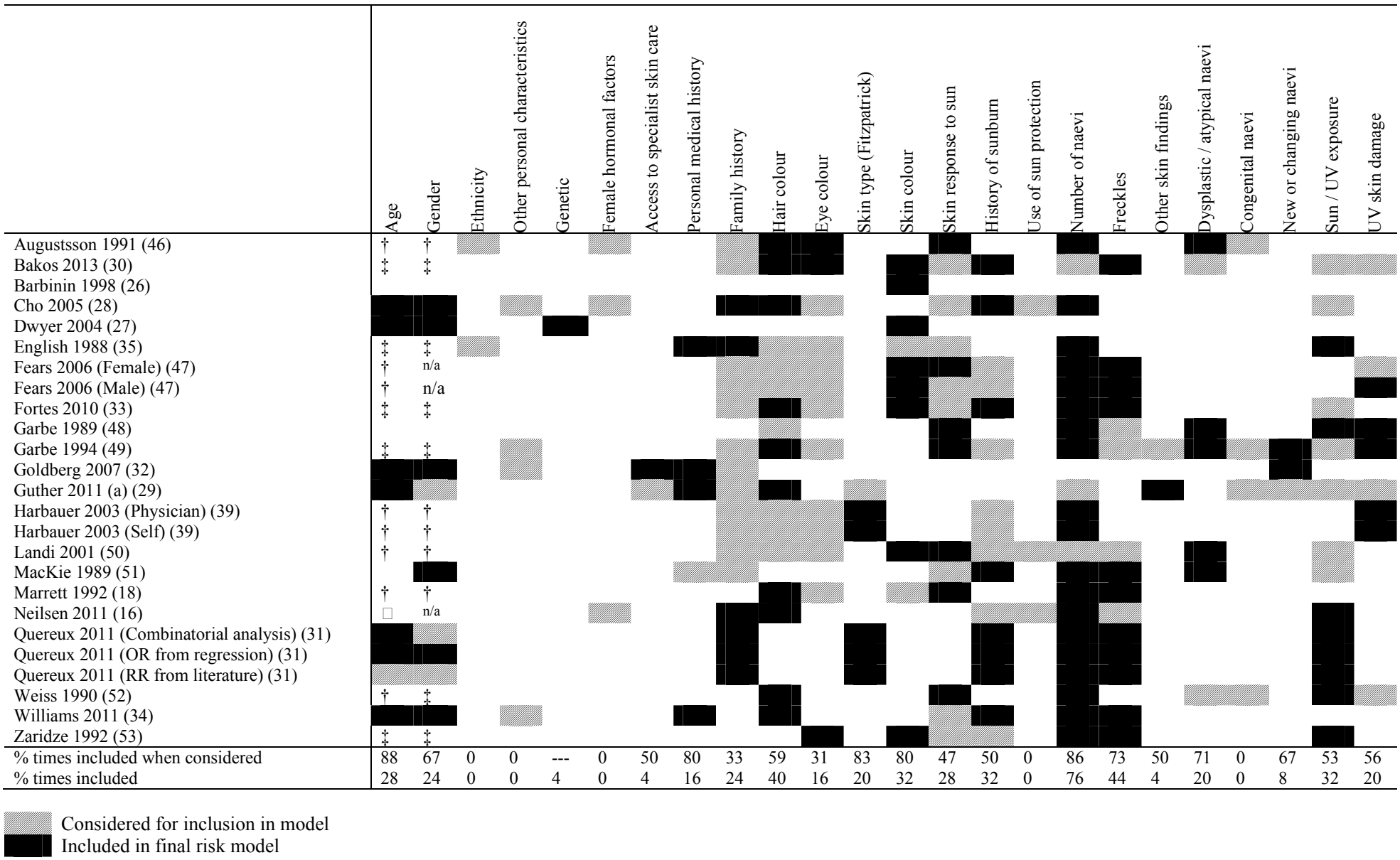
(d) The Hosmer-Lemeshow statistic assesses whether or not the observed event rates match expected event rates in subgroups of deciles of fitted risk values. A non-significant p value indicates a well calibrated model.
AUROC – area under the receiver operator curve; Sens – sensitivity; Spec – specificity; PPV – positive predictive value; NPV – negative predictive value

Table 2. Details of validation studies

| Risk model | Study | Country, Year | Study design | Data collection method | Selection of cases | Selection of controls | Number of cases:controls (Participation rate, %) | Discrimination | Accuracy | Utility |
|---|---|---|---|---|---|---|---|---|---|---|
| Fortes 2010 (33) | Fortes 2010 (33) | Brazil, 2005-8 | Case Control | Interview administered questionnaire and examination | Caucasian individuals with histologically confirmed primary melanoma, > 18 and resident in study area | Caucasian patients from general wards without a personal history of skin cancer matched by age and sex | 64 (97%): 53 (100%) | AUROC 0.79 (0.70-0.86) | Cut-off level ≥ 3: Sens 79.6; Spec 60 | |
| Williams 2011 (34) | Williams 2011 (34) | USA, 1997 | Case Control (subset of original study) | Telephone survey | Patients with primary invasive cutaneous melanoma from surveillance epidemiology and cancer register | Random digit dialling | 25% of 386 (80%): 727 (63%) (a) | AUROC 0.70 (0.64-0.77) | Cut off 25: Sens 61; Spec 80<br>Cut off 28: Sens 50; Spec 85<br>Cut off 30: Sens 42; Spec 90<br>Cut off 34: Sens 29; Spec 95 (b) | |
| MacKie 1989 (51) | Jackson 1998 (54) | UK, 1995 | Cohort | Self-completion of questionnaire and examination | Consecutive patients > 16 visiting their doctor | | 388 (26%) (c) | | | Agreement of self-report and clinical examination: κ 0.43 - 0.67 |
| Neilsen 2011 (16) | Westerdalh 1996 (55) | Sweden, 1990-4 | Cohort | Postal questionnaire | Random sampling of women who had responded to initial questionnaire 1-3 years previously | | 670 (84%) | | | Test-retest reliability of questionnaire: κ 0.54 – 0.83 |
| Quereux 2011 (1, 2, 3) (31) | Quereux 2010 (56) | France, 2006-7 | Cohort | Self-completion of questionnaire and examination | Consecutive patients 18-70 years visiting their doctor | | 1358 (100%) (d) | | | Agreement of self-report and clinical examination: % correct answers 79.9 – 98.1 |
| Quereux 2011 (3) (31) | Quereux 2012 (57) | France, 2009 | Cohort | Self-completion of questionnaire and examination of patients at high risk | Consecutive patients > 18 visiting their doctor | | 1039 (43%) (e) | | | Efficiency 11.54 (p=0.0016) (f) |

(a) Of 1751 who agreed to take part, 1024 (58%) were subsequently excluded as they were not eligible
(b) Based on both development and test populations
(c) The initial response rate to the questionnaire was 66%. Of those who responded, 388 (26%) attended for a skin examination
(d) Of 1500 patients agreeing to take part,42 (2.8%) were excluded as they were not eligible and 100 (6.7%) for incomplete data
(e) 7953 completed the questionnaire whilst visiting their GPs. 2404 were high risk and 1039 (43%) of those consulted a dermatologist. Of those 95 had a biopsy and a melanoma was found in 10
(f) The interpretation of this is that to detect a new case of melanoma it is necessary to screen 11.54 times fewer patients than with non-targeted screening

Table 3. Factors considered and then included in final risk models



(a) Skin type unclear from methods - use I-VI but no further details
† Model adjusted for age and/or gender
‡ Case-control study matched by age and/or gender
 Did provide results for age stratification in two age ranges
(b) Computed only for factors considered in more than one model

**Figure 1**

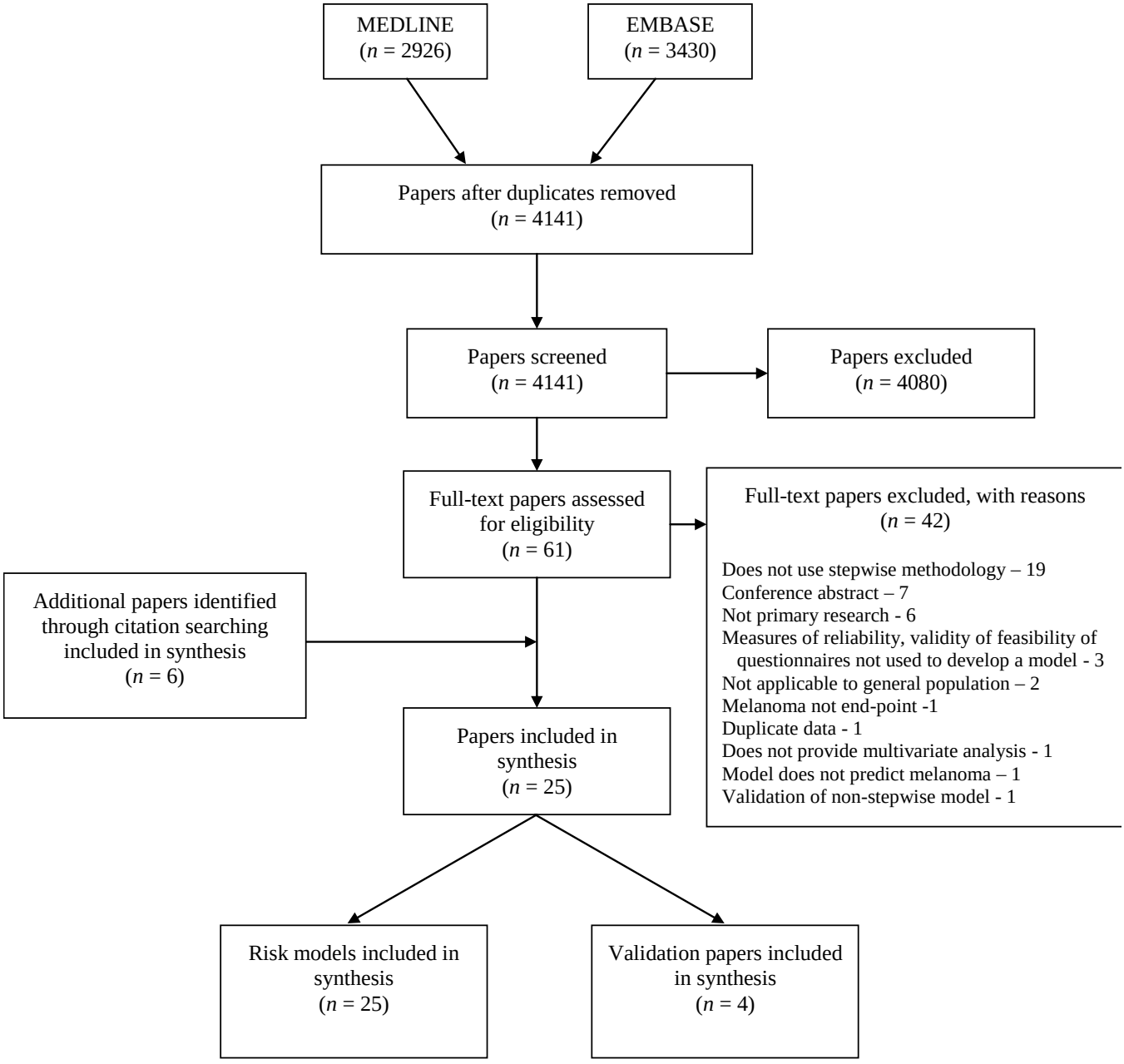


Figure 2