Article

Using the Prediction Model Risk of Bias Assessment Tool (PROBAST) to Evaluate Melanoma Prediction Studies

Isabelle Kaiser [1], Sonja Mathes [2], Annette B. Pfahlberg [1], Wolfgang Uter [1], Carola Berking [3], Markus V. Heppt [3,4], Theresa Steeb [3], Katharina Diehl [1] and Olaf Gefeller [1*]

[1] Department of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany; isabelle.kaiser@fau.de (I.K.); annette.pfahlberg@fau.de (A.B.P.); wolfgang.uter@fau.de (W.U.); katharina.diehl@fau.de (K.D.)
[2] Department of Dermatology and Allergy, Technische Universität München, 80802 München, Germany; sonja.mathes@mri.tum.de"
[3] Department of Dermatology, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91054 Erlangen, Germany; carola.berking@uk-erlangen.de (C.B.); markus.heppt@uk-erlangen.de (M.V.H.); theresa.steeb@uk-erlangen.de (T.S.)
[4] Comprehensive Cancer Center Erlangen-European Metropolitan Area of Nuremberg (CCC ER-EMN), 91054 Erlangen, Germany
* Correspondence: olaf.gefeller@fau.de; Tel.: +49-9131-85-22750

Simple Summary: The rising incidence of cutaneous melanoma over the past decades, combined with a general interest in cancer risk prediction, has led to a high number of published melanoma risk prediction models. The aim of our work was to assess the validity of these models in order to discuss the current state of knowledge about how to predict incident cutaneous melanoma. To assess the risk of bias, we used a standardized procedure based on PROBAST (Prediction model Risk Of Bias ASsessment Tool). Only one of the 42 studies identified was rated as having a low risk of bias. However, it was encouraging to observe a recent reduction of problematic statistical methods used in the analyses. Nevertheless, the evidence base of high-quality studies that can be used to draw conclusions on the prediction of incident cutaneous melanoma is currently much weaker than the high number of studies on this topic would suggest.

Abstract: Rising incidences of cutaneous melanoma have fueled the development of statistical models that predict the individual melanoma risk. Our aim was to assess the validity of published prediction models for incident cutaneous melanoma using a standardized procedure based on PROBAST (Prediction model Risk Of Bias ASsessment Tool). We included studies that were identified by a recent systematic review and updated the literature search to ensure that our PROBAST rating included all relevant studies. Six reviewers assessed the risk of bias (ROB) for each study using the published "PROBAST Assessment Form" that consists of four domains and an overall rating of ROB. We further examined a temporal effect regarding changes in overall and domain-specific ROB rating distributions. Altogether 42 studies were assessed, of which a vast majority (n=34; 81%) was rated as having high ROB. Only one study was judged as having low ROB. The main reasons for high ROB ratings were the use of hospital controls in case-control studies and the omission of any validation of prediction models. However, our results of the temporal analysis showed a significant reduction in the number of studies with high ROB for the domain analysis. Nevertheless, the evidence base of high-quality studies that can be used to draw conclusions on the prediction of incident cutaneous melanoma is currently much weaker than the high number of studies on this topic would suggest.

Keywords: risk prediction; prediction models; risk of bias; PROBAST; melanoma

1. Introduction

 doi:10.20944/preprints202205.0091.v1

Cutaneous melanoma is one of the most lethal forms of skin cancer that accounts for the majority of skin cancer deaths [1]. The incidence rates of melanoma have been growing dramatically over the past decades in most fair-skinned populations worldwide with annual increases of 3 to 7% [2-4]. The highest incidence rates by far are observed in Australia and New Zealand [5], although the incidence rates in these two countries are now stabilizing or even slightly declining following intensive preventive efforts [4,6]. Other regions with high melanoma incidences and ongoing rising trends are Western and Northern Europe, as well as North America [2,4,5]. The increasing incidence rates over the past decades, a better understanding of genetic and environmental risk factors, and a growing general interest in cancer risk prediction have fueled the development of risk prediction models for melanoma. Risk prediction models enable the proper identification of individuals at high risk of developing the disease. They are essential tools for more effective, targeted screenings of individuals at higher risk as a part of secondary prevention strategies.

Although a variety of prediction models for assessing the individual melanoma risk have been published over the past 40 years, none has become widely accepted in clinical practice. An essential prerequisite for a reliable risk prediction model, that can be implemented in clinical practice, is a properly conducted, well-reported and validated development study. Currently, many risk prediction models are not externally validated [7-9], which means that the performance of the model has not been evaluated in an independent dataset. This is important, because shortcomings in study design, methods, conduct, or analysis often lead to overoptimistic predictive performance estimates of the model in the development study [10]. This overoptimism, i.e. the overestimation of the model's predictive ability, results typically from an overfitting of the developed model to specific characteristics of the dataset that was used to develop the model. When the prediction model is applied to new data, the predictive performance is worse than before [11,12]. This in turn can result in inaccurate models leading to false predictions, which would be detrimental when using the model in clinical practice for risk stratification. False predictions may lead to either unnecessary or insufficient interventions that may influence the health of those affected by the wrong prediction. Thus, it is necessary to evaluate the presence of systematic error in risk prediction studies which may jeopardize the validity of conclusions drawn from such studies. Regarding the assessment of bias in melanoma risk prediction, there is still a need to catch up with other areas of prediction modeling. None of the existing systematic reviews on melanoma prediction studies included a risk of bias (ROB) assessment, which motivated us to fill this gap using the recently developed PROBAST (Prediction model Risk Of Bias ASsessment Tool; https://www.probast.org) methodology [13].

PROBAST was developed in 2019 to facilitate the tailored ROB assessment for studies exploring prediction models. It provides a methodological quality assessment of primary studies that report on the development, validation or update of prediction models. The tool can be used for all clinical domains, predictors, outcomes, and modelling techniques [13,14].

The primary objective of this work was to assess the validity of published prediction models for incident cutaneous melanoma using a standardized procedure based on PROBAST. In addition to describing the PROBAST results for the overall and domain-specific ratings, we discuss the consequences of our assessment results on the current state of knowledge about how to predict incident cutaneous melanoma.

2. Materials and Methods

2.1. Study Selection and Eligibility Criteria

Details on the study selection and eligibility criteria have been published before in a report describing the reporting quality of melanoma prediction studies [15]. In brief, we included studies reporting the development and validation of models for predicting the

 doi:10.20944/preprints202205.0091.v1

individual risk of occurrence of cutaneous melanoma. Only studies providing either absolute risks or risk scores, or report mutually adjusted relative risks for primary cutaneous melanoma were eligible. The set of studies to be assessed was based on a recent systematic review on melanoma prediction modeling [7] that updated two earlier systematic reviews on this topic [8,9]. To ensure that our PROBAST rating included all relevant studies, we performed a literature update for the time interval since the end of the search period for the systematic review [7], that is, February 2020 and August 2021. In particular, the forward snowballing technique [16] was applied to all three systematic reviews [7-9] and an electronic literature search in PubMed using the same search string as in [7] was conducted.

2.2. PROBAST Rating

The ROB of each study was assessed independently by six reviewers (I.K., S.M., M.V.H., T.S., K.D., O.G.). The reviewer panel was multidisciplinary and consisted of reviewers with methodological (I.K., O.G.), clinical (S.M., M.V.H.), and public health (T.S., K.D.) backgrounds at different levels of experience. All reviewers used the PROBAST tool provided on https://www.probast.org/. Furthermore, a web-based input tool was created for data collection using the software SoSci Survey version 3.2.21 (SoSci Survey GmbH, Munich, Germany) [17]. All six reviewers assessed all 42 studies. Disagreements between the reviewers regarding the ROB rating were resolved in ten virtual consensus meetings. In case of sustained disagreements, two independent referees (A.B.P., W.U.) decided.

The PROBAST tool consists of the four domains participants, predictors, outcome, and analysis. For each domain, the ROB was rated individually as either low, high, or unclear. Several signaling questions, that were answered as yes, probably yes, no, probably no, or no information, assisted the reviewer in judging the ROB for each domain. Finally, an overall ROB was assigned to the study based on the ratings in the four domains. According to the given rules in the PROBAST tool [13], the overall ROB is obtained by taking the lowest rating of any domain-specific ROB ("worst score counts principle"). Consequently, the overall ROB was high if at least one of the four domains was rated as high. If at least one domain was judged as unclear and all other domains as low, the overall ROB was unclear. Thus, a study only received a low overall ROB if all four domains were judged as having low ROB. However, according to PROBAST guidance, downgrading to high or unclear ROB should be considered, if a prediction model was developed without any external validation. In the absence of external validation, the model evaluation was only considered as low ROB, if the development was based on a very large data set and included some form of internal validation.

Since the ROB rating strongly depends on the reviewer's judgment, some decision rules for the specific setting of melanoma prediction studies were defined by the reviewers in advance to establish a common standard for the rating (see section 2.3). The decision rules overruled individual ratings and referee decisions. Therefore, all ratings were checked for consistency with the self-defined decision rules and discussed in case of disagreement.

2.3. Description of Domains and Decision Rules

Domain 1: Participants

This domain was related to possible sources of bias associated with the data sources and participant selection. In general, the selection of participants should represent the target population [14]. We defined the following specific rules for this domain: A study received a high ROB if 1) in case-control studies, the cases were recruited in a single center or the controls consisted of hospital controls, 2) in cohort studies, no population sample was used or the study population was self-selected, or 3) in studies based on risk estimates from meta-analyses, the studies included in the meta-analyses met the criteria for a high ROB in this domain. If the references of the studies included in the meta-analyses were not given, the ROB is rated as unclear.

 doi:10.20944/preprints202205.0091.v1

Domain 2: Predictors
The domain predictors covered possible sources of bias related to the selection and assessment of predictors. The risk factors had to be defined and collected in the same way for all study participants [14]. Our specific decision rules included that pooled studies and meta-analyses were rated with a high ROB as default, as heterogeneity in definition and assessment of predictors between the included studies was assumed. If it was explicitly described that no heterogeneity existed, e.g., by using identical protocols for the risk factor assessment, a low ROB rating was possible. Furthermore, the use of risk factors with possible recall bias in case-control studies led to an unclear ROB rating. These included predictors related to natural (solar) and artificial UV exposure in the past.

Domain 3: Outcome
The third domain covered a possible bias generated by the definition or determination of the outcome. Objective outcomes, such as histologically confirmed diagnoses, are less susceptible to bias than outcomes that require subjective interpretation or are based on participants' self-assessment [14]. Consequently, we specified the following rule for ROB ratings: outcomes without verified melanoma diagnosis, e.g., self-reported lifetime melanomas that were assessed via questionnaire, are rated as high ROB.

Domain 4: Analysis
The focus of the last domain was a potential bias in the estimated predictive performance triggered by inappropriate analysis methods or omission of important statistical considerations. Aspects of the analysis to be considered for the bias rating included: 1) whether the sample size was sufficient, 2) whether predictors were incorporated appropriately into the model, 3) whether missing data were handled adequately, 4) whether the predictive performance of the model was evaluated systematically and 5) whether model overfitting was accounted for [14]. We defined the lack of internal and external validation as a sufficient criterion for a high ROB. Another criterion for a high ROB rating was the lack of quantitative information about performance measures. Thus, at least one performance measure and one form of validation had to be reported to obtain a low ROB, provided that the analysis regarding the other aspects was sound. If the analysis contained components whose effect on the results was unclear or the description allowed no definite categorization as either low or high ROB, the domain received an unclear ROB rating.

For all domains, if the information given in the study publications was too limited to allow an assessment of ROB, the respective domain was rated unclear. The full list of specific decision rules for high and unclear ROB that was updated after the rating and consensus meetings can be found in the Supplement.

2.4. Statistical Analysis

The results of the ROB assessment were analyzed descriptively and presented as absolute and relative frequencies. Group comparisons between studies reporting solely model development and studies reporting both model development and external validation were done by the Fisher's exact test in its modified version for 2x3-tables. A possible temporal effect regarding changes in overall and domain-specific ROB rating distributions was additionally investigated. To this end, the studies were divided into three groups based on their year of publication. Using the tertiles of the distribution of publication years we defined the following three time intervals: "1988-2006" (n=14), "2007-2014" (n=15) and "2015-2021" (n=13). We used the Mantel test [18] to check for an association between ROB ratings and time interval as the Mantel test incorporated the ordinal structure of both variables which the Chi-squared test, the statistical standard test in this situation, would have ignored. Due to the sparse data situation we were facing in our study we employed the exact version of the Mantel test based on the network algorithm developed by Mehta and Patel [19]. P-values smaller than 0.05 were interpreted as indicating statistical significance.

 doi:10.20944/preprints202205.0091.v1

All statistical analyses were performed using the R software version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria) [20].

3. Results

3.1. Study Characteristics

In total, we included 42 studies in our PROBAST rating. Forty studies [21-60] were adopted from the most recent systematic review about risk prediction models for melanoma that was published in 2020. The remaining two recent studies [61,62] were identified through the updated literature search. Study characteristics are summarized in Table A1 in the Appendix. Thirty-five of the 42 studies (83%) solely described the development of a melanoma risk prediction model, while seven studies (17%) reported both development and external validation. The publication years of the studies ranged from 1988 to 2021, with a pronounced increase in the number of studies in the last decade of this interval. The majority of studies were case-control studies (n=30). Ten studies used a cohort study design and two studies used published material from meta-analyses to determine predictors and risk estimates.

3.2. Results of Risk of Bias Rating

Results of the domain-specific and overall ROB ratings of our set of 42 studies are shown in Figure 1. The individual ROB ratings of the 42 studies are included in Table A1. In the following, the results for the individual domains are described.

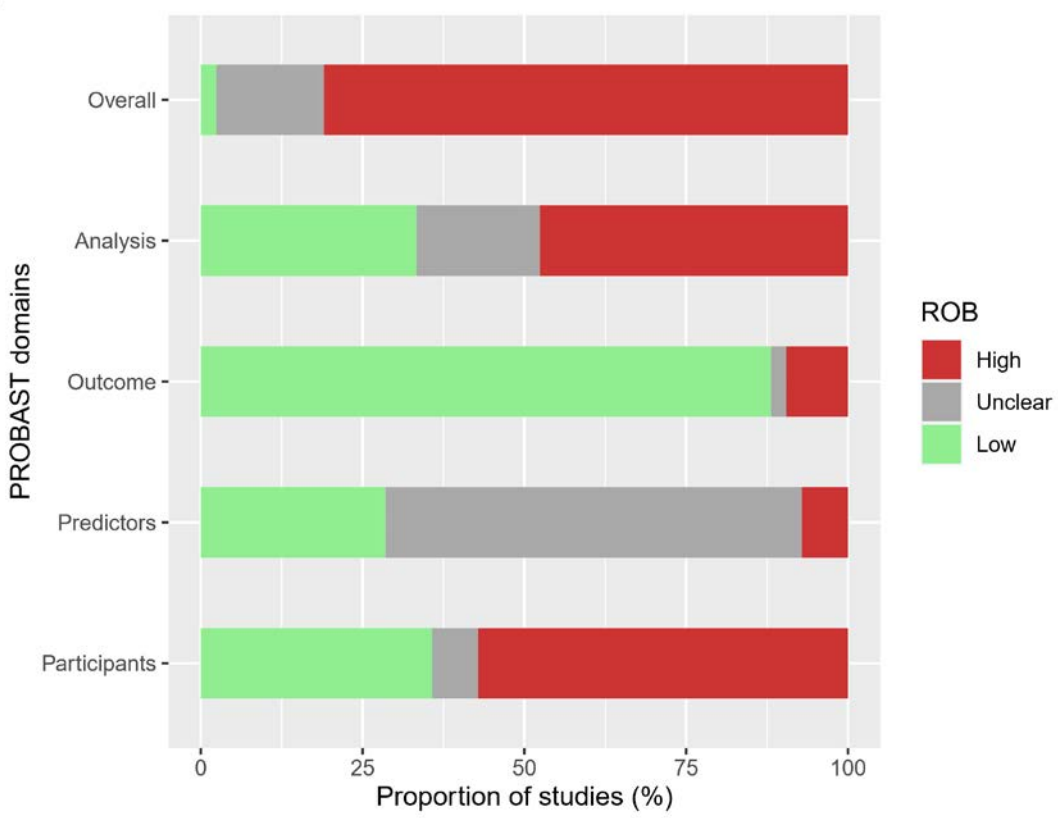

Figure 1: Risk of bias rating overall and per domain (N=42 studies).

Domain 1: Participants

In the participants domain 24 studies (57%) were rated as high, 3 studies (7%) as unclear and 15 studies (36%) as low ROB (see Figure 1). In 15 studies, the selection of controls

 doi:10.20944/preprints202205.0091.v1

in case-control study designs was decisive for the high ROB rating, mainly because of the use of hospital controls (n=14). In addition, four studies based on meta-analyses received a high ROB as they each contained studies with a high ROB. In four cohort studies, the use of a self-selected screening population resulted in a high ROB rating. Further reasons that led to an unclear or high ROB rating are listed in Table 1.

Table 1: Reasons for unclear (N=3) and high (N=24) ROB ratings in the participants domain

Unclear ROB		**High ROB**	
Reason	n (%)	Reason	n (%)
Limited information	2 (67%)	Hospital controls (case-control studies)	14 (58%)
Data from a costumer data-base offering genetic analyses without information regarding population coverage	1 (33%)	Meta-analysis including studies with high ROB	4 (17%)
		Self-selected screening population/ no population sample (cohorts)	4 (17%)
		Highly selected sample	1 (4%)
		Mixed bag of controls (including hospital controls)	1 (4%)

Domain 2: Predictors

Three studies (7%) were rated as high ROB in the predictors domain due to heterogenous predictor assessment of studies included in the meta-analyses or pooled studies (Figure 1, Table 2). Furthermore, 27 studies (64%) were rated as unclear. In the majority of cases (n=21) the reason was potential recall bias in case-control studies due to predictors related to UV-exposure in the past. Three studies did not provide enough information for the evaluation of potential bias which also lead to an unclear ROB rating. The remaining three studies with an unclear ROB rating in the predictors domain suffered from discrepancies between development and validation data sets. Twelve (29%) of the included studies were rated as low ROB.

Table 2: Reasons for unclear (N=27) and high (N=3) ROB ratings in the predictors domain

Unclear ROB		**High ROB**	
Reason	n (%)	Reason	n (%)
Potential recall bias	21 (78%)	Pooled study or meta-analysis with heterogenous predictor assessment	3 (100%)
Limited information	3 (11%)		
Replacement of predictors in validation	1 (4%)		
Unclear harmonization of predictor variables in development and validation datasets	1 (4%)		
Missing predictors in validation dataset	1 (4%)		

Domain 3: Outcome

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 May 2022 doi:10.20944/preprints202205.0091.v1

The outcome domain comprised the highest proportion (n=37, 88%) of low ROB ratings among all four domains in our investigation. The ROB of one study (2%) was rated as unclear due to limited information regarding the definition and assessment of the outcome (Figure 1, Table 3). Four studies (10%) received a high ROB rating. Three of the four studies did not use verified outcomes: self-reported lifetime melanomas (n=2) or suspected melanomas (n=1). The fourth study used a composite outcome consisting of melanoma and cannot-exclude-melanoma/severely dysplastic nevi.

Table 3: Reasons for unclear (N=1) and high (N=4) ROB ratings in the outcome domain

Unclear ROB		**High ROB**	
Reason	n (%)	Reason	n (%)
Limited information	1 (100%)	Self-reported outcome	2 (50%)
		Composite outcome (melanoma and severely dysplastic naevus)	1 (25%)
		Suspected melanoma as outcome	1 (25%)

Domain 4: Analysis

In the analysis domain, eight studies (19%) had an unclear ROB, whereas for 20 studies (48%) the ROB was rated as high and for 14 studies (33%) as low (Figure 1). Reasons for an unclear ROB rating were, e.g., limited information regarding the analysis (n=4) and non-standard handling of predictors during the statistical analysis entailing unknown impact on the results (n=2), see Table 4. The main reason for high ROB was a missing internal and external validation (n=19). In several cases multiple reasons for a single study led to a high ROB rating. However, in Table 4 we have listed only the reasons that were decisive for our rating, which was primarily the lack of validation. The lack of internal and external validation often occurred in combination with missing performance measures (n=12), a limited sample size (n=3) and/or missing information regarding one or multiple aspects of the analysis (n=14).

Table 4: Reasons for unclear (N=8) and high (N=20) ROB ratings in the analysis domain

Unclear ROB		**High ROB**	
Reason	n (%)	Reason	n (%)
Limited information	4 (50%)	No validation	19 (95%)
Non-standard handling of predictors during the analysis	2 (25%)	Limited sample size	1 (5%)
Rounding of model coefficients to define the risk score	1 (12.5%)		
Several aspects of analysis unclear	1 (12.5%)		

Overall ROB

Overall, only one study (2%) received a low ROB rating, whereas 7 studies (17%) were judged to have an unclear ROB. Four [26,27,57,58] of these seven studies received their unclear ROB rating due to an unclear ROB rating in a single domain, while the remaining three studies [30,50,60] had an unclear ROB rating in two domains. The majority of studies (n=34; 81%) were associated with a high ROB (Figure 1). For one study [54] we used the option of downgrading according to PROBAST guidance. The study received a low ROB rating in the domains participants, outcome and analysis, and an unclear rating in the predictors domain that would have resulted in an overall unclear ROB accordingly. However, due to its small sample size and lacking external validation the study was downgraded to high ROB.

 doi:10.20944/preprints202205.0091.v1

3.3. Risk of Bias Rating Stratified by Study Type

A comparison of the overall ROB ratings by study type showed that development-only studies received a high ROB rating more frequently than studies that both developed and validated their model (89% vs. 43%, p=0.017). While in the domain analysis 80% of the development studies got an either unclear or high ROB rating, all development and validation studies received a low ROB rating (p<0.001) for this domain. Further, the proportions of studies with high ROB scores in the predictors and participants domains were lower for studies including an external validation than for development-only studies (0% vs. 9% and 43% vs. 60%, respectively), but these differences missed statistical significance.

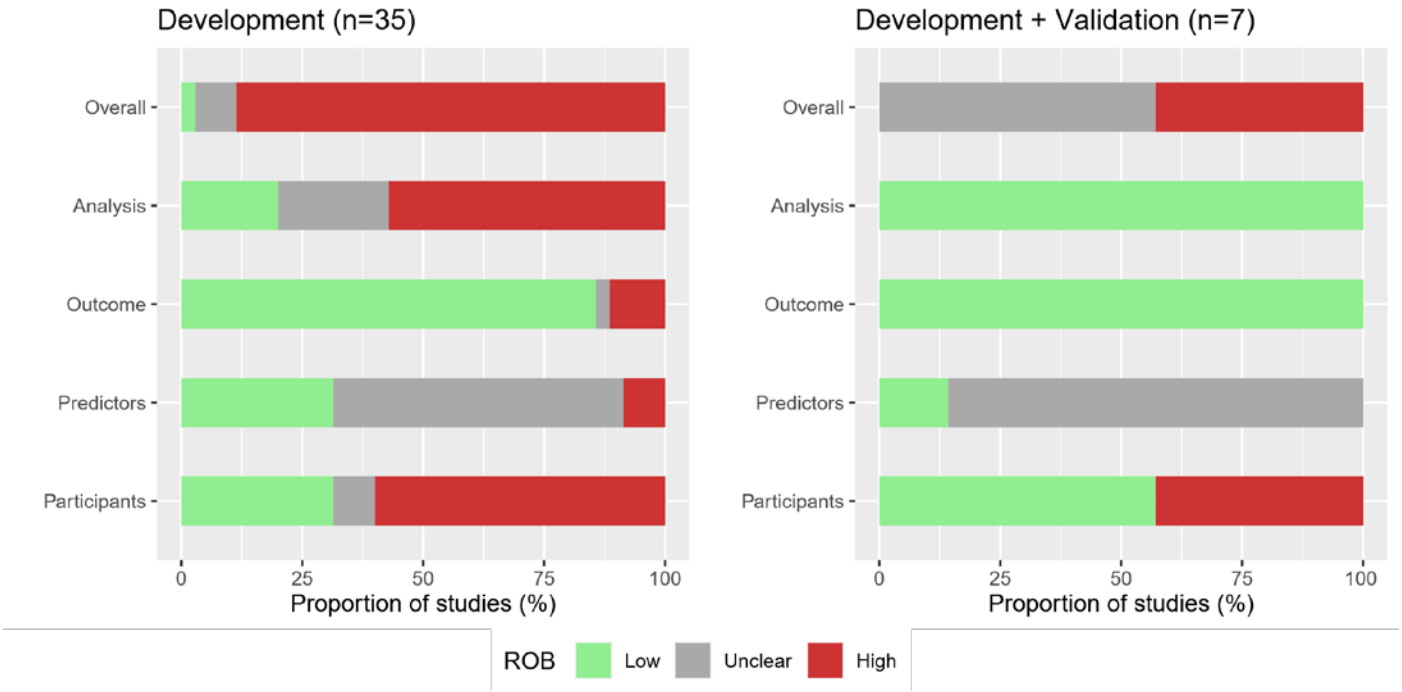

Figure 2: ROB rating in development studies (N=35) and development and validation studies (N=7).

3.4. Temporal Analysis

The proportion of studies with low, unclear and high ROB ratings in the three time intervals is visualized in Figure 3. For the domain analysis we found a clear temporal trend of better ROB ratings for more recent studies. The proportion of studies rated as high ROB decreased significantly over the three time intervals (79% vs. 40% vs. 23%, p=0.004). For the three other domains we did not observed such clear-cut temporal development of ROB rating distributions. The overall ROB rating distributions in the three time intervals indicated some improvement: the proportion of studies rated as high ROB decreased steadily from 93% in 1988-2006 over 80% in 2007-2014 to 69% in 2015-2021, but this decline missed statistical significance (p=0.10).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 May 2022 doi:10.20944/preprints202205.0091.v1

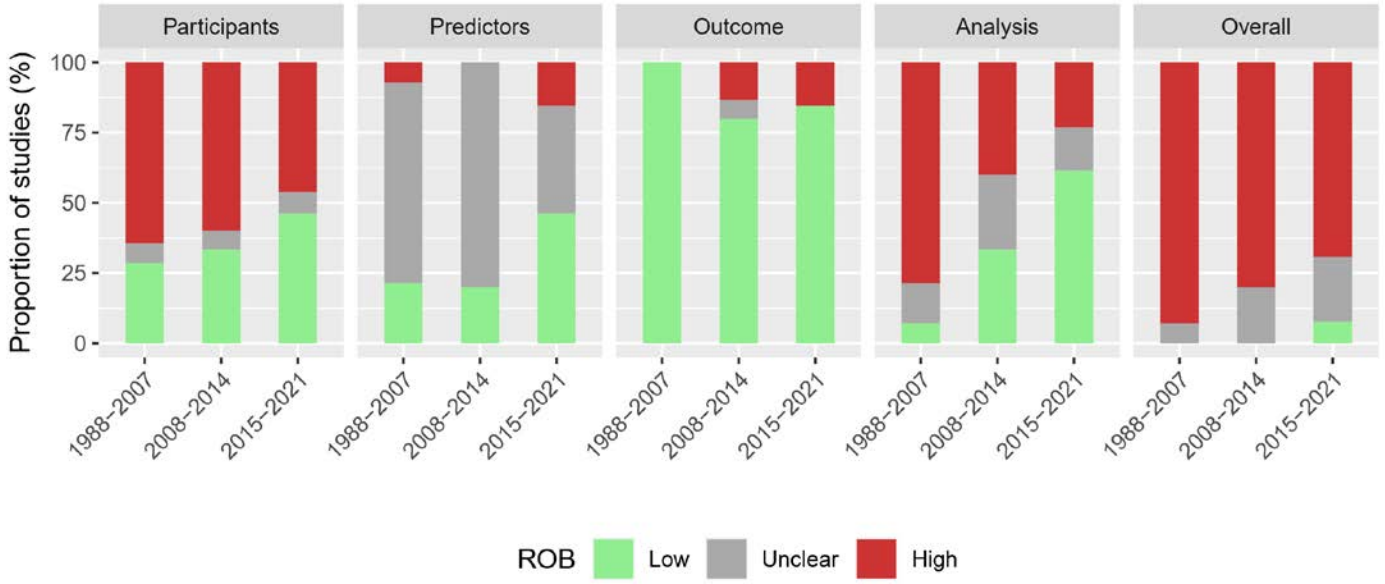

Figure 3: Comparison of proportions of studies with high, unclear and low ROB for domain-specific and overall ratings in the three time intervals "1988-2006" (N=14),"2007-2014" (N=15) and "2015-2021" (N=13).

4. Discussion

The results of our ROB assessment showed a clear deficit of valid risk models for melanoma prediction, as the vast majority (81%) of the included 42 studies was associated with a high ROB. Thus, the evidence base of high-quality studies that can be used to draw conclusions on the prediction of incident cutaneous melanoma is currently much weaker than the high number of studies on this topic would suggest.

Only one [49] of the 42 studies had a low overall ROB score. The study was the QSkin Sun and Health Study, a prospective cohort study of 43,794 participants randomly sampled from the population of Queensland, Australia in 2011 [63]. Up to now, the QSkin study is the largest prospective study ever conducted specifically to address melanoma and other skin cancer outcomes. The study report from 2018 [49] described separately the prediction of invasive and any melanoma (incl. in situ melanoma) using self-assessed risk factors. The model for predicting invasive melanoma included the following seven risk factors: age, sex, tanning ability, number of nevi at age 21 years, hair color, number of actinic skin lesions destroyed, and sunscreen use when outdoors during the past year. The same risk factors were also part of the prediction model for any melanoma that additionally included five risk factors, e.g., family history of melanoma and number of skin checks by a doctor during the past three years. Although the study raised no concerns regarding systematic error in study design, conduct, methods, and analysis, the application of its risk models in clinical practice is limited by their moderate predictive performances: The model discrimination, as described by the C-index, was only 0.69 (95%-CI: 0.62, 0.76) for the invasive melanoma model and 0.72 (95%-CI: 0.69, 0.75) for the any melanoma model showing that additional explanatory variables are required to improve the predictive performance.

Furthermore, four publications [26,27,57,58] had overall an unclear ROB score resulting from a domain-specific unclear ROB rating in a single domain (in all four cases the domain predictors). These publications described externally validated models from the same population-based case-control study. In all four publications data from the Australian Melanoma Family Study [64] were used to develop the prediction model. This study only included cases diagnosed with invasive cutaneous melanoma at age 18-39 years and is therefore highly selective. Data from the Leeds Melanoma Case-Control Study [65] were

 doi:10.20944/preprints202205.0091.v1

used to validate the model (in [58] data from three additional case-control studies served for additional external validations). Two of the publications incorporated [26,27] genotype information, while the remaining two [57,58] focused on non-genetic risk factors. The difference between the two non-genetic prediction models related to the inclusion of only self-reported risk factors in [58] and the use of physician-assessed risk factors related to skin phenotype in [57]. The models differed considerably in their performance, the AUC describing model discrimination ranged from 0.66 (95%-CI: 0.63, 0.68) for the model including only self-assessed risk factors without genotype information [58] to 0.79 (95%-CI: 0.76, 0.81) for the model including physician-assessed risk factors and genotype information related to the MC1R genotype [27]. The main driver of the increments in the AUC was the incorporation of physician-assessed nevi counts instead of self-assessed nevi density. The use of genotype information had only a moderate impact, contrary to what one would expect from the increasing popularity of genetic risk factors in recent years.

The selection of risk factors has not only a significant impact on the performance of the model, but is also related to possible bias, especially in case-control studies. The high proportion of studies with an unclear ROB rating in the predictors domain resulted primarily from the use of predictors related to past UV exposure. Whenever such predictors are ascertained in retrospective case-control studies, estimation of their impact on melanoma risk is prone to recall bias, i.e. a special form of exposure misclassification in case-control studies. For melanoma, the presence of recall bias has attracted considerable attention and has been analyzed using different approaches in various studies [66-72]. There has been no clear conclusion regarding the magnitude of the bias [73,74]. The consequences of incorporating such predictors into melanoma prediction models have not been discussed by any of the developers of these models and remain unclear. Another source of bias in case-control studies that led to most high ROB ratings in the participants domain is the use of hospital controls. In order to prevent bias in case-control studies, the controls must be selected independent of exposure and need to represent the study population at risk of becoming cases [75]. Although the selection of hospital controls has some practical advantages, e.g., they are readily accessible and usually cooperative, the presence of unsuspected associations between the reason for hospital visit and the factors of interest can lead to systematically distorted estimates [76-78]. Hospital controls are likely to have a higher frequency of hazardous exposures compared to the general population [79].

The large numbers of high and unclear ROB ratings demonstrate the need to reduce bias in future studies. One possibility is to consider the criteria of ROB tools already in the study planning stage. Thus, sources of bias related to the selection of the study population and the definition of outcome assessment, for example, could be avoided. Another opportunity for reducing bias can be found in the analysis domain. The main reason for high ROB ratings was the lack of validation (internal or external), often combined with missing evaluation of model performance. However, we have seen a positive temporal trend in this domain: The proportion of high ROB ratings has significantly decreased by more than 50%. This development shows that the journals have been more rigorous in applying pertinent quality standards in recent years, particularly concerning the methodology employed during statistical analysis. An important additional contribution to the positive development is made by the large number of checklists and accompanying guidance papers that have been published in recent years. These include reporting guidelines such as TRIPOD (Transparent Reporting of a multivariate prediction model for Individual Prognosis Or Diagnosis) [80], which provides a checklist of 22 items essential for transparent reporting of a prediction model study [15]. It ensures that all relevant key details on the development process and model performance, that are needed to objectively appraise the validity and usefulness of the model, are reported. Furthermore, guidelines that directly include a ROB tool, such as the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [81] for systematic reviews and meta-analyses. PRISMA is already required by many scientific journals, which has demonstrably improved the conduct and reporting of systematic reviews and meta-analyses [82]. Other

 doi:10.20944/preprints202205.0091.v1

tools for the assessment of ROB are, e.g., the Cochrane ROB tool [83] for randomized controlled trials, which was published in 2011 and updated in 2019 [84]. All of these have the potential to ensure a high transparent quality of studies developing risk prediction models if applied properly. However, we conclude from our results that in order to better implement and advance knowledge about melanoma risk prediction, it is essential to expand the application of existing guidelines in practice to improve the quality of prediction model studies, especially regarding study design and standardization of methodology to conduct this type of studies.

To the best of our knowledge, this is the first assessment of bias in melanoma prediction studies, hence there is no direct comparison of our results with other papers. However, comparisons with ROB results from assessments in other clinical domains are possible. The two systematic reviews by Sassano et al. [85] and Su et al. [86] addressing risk prediction of colorectal cancer and caries, respectively, involved ROB assessment with PROBAST. Both criticized an insufficient number of high-quality studies in their clinical domains, the proportion of studies with high ROB being 94% and 78%, respectively. In 2021 a meta-review by de Jong et al. [87] including 50 systematic reviews across various clinical domains that all used PROBAST for ROB assessment was published. The ROB rating from a total of 1510 individual studies was reported. Similar to our results, the authors observed predominantly unclear and high ROB ratings at the domain-specific levels, while results of the overall ROB were not reported. The domain analysis showed with 69% the highest proportion of high ROBs, which is higher than in our rating where the proportion of high ROBs in this domain was 48%. Unlike ours, the results were stable over time. This shows that the positive temporal trend towards higher quality standards concerning statistical methodology, which is visible in melanoma prediction studies, has not yet reached all clinical domains.

During our assessment, we encountered some obstacles in the practical application of PROBAST, which show that the tool is not easily applicable in all situations. According to PROBAST, case-control studies do not represent appropriate data sources and should be rated with high ROB as default. Though case-control studies are more prone to bias, this is not primarily due to the study design itself but due to practical problems in study conduct, some of which have already been described above. Per se, case-control studies can yield results as valid as cohort studies, if they are properly planned, conducted and analyzed [88]. In addition, some signaling questions that should support the ROB rating, such as the questions "Was the outcome determined without knowledge of predictor information?" and "Was the time interval between predictor assessment and outcome determination appropriate?" in the outcome domain, are only applicable for prospective studies. In case-control studies the outcome status is already known when the participants are being selected and thus before the predictor assessment. In general, the continuous adaptation and improvement of rating tools is necessary to further increase their applicability and popularity. In particular, the PROBAST tool should therefore be amended or supplemented for study design-specific features to ensure unequivocal assessment. Otherwise, systematic reviews employing PROBAST need to redefine generic signaling questions for their application.

Due to above-mentioned obstacles in the applicability of the tool to case-control studies, which accounted for 71% of our included studies, but also to provide a consistent basis for our rating, we defined some specific decision rules that overruled the decisions of individual raters and those of the referees. Since the decision rules were designed to the best of our knowledge but were not validated separately, this may have resulted in some bias in our ROB ratings and constitutes a limitation of our work. Additionally, the ROB judgement is subjective and does not lend itself to a clear objective rating. As different raters may have come to different conclusions on how to rate the individual PROBAST domains, it cannot be ruled out that another group of raters would have come to other results regarding the PROBAST ratings in the same set of melanoma prediction studies. We have

 doi:10.20944/preprints202205.0091.v1

tried to minimize this rater dependence by defining the decision rules, by holding consensus meetings to resolve discrepancies in ratings, and by involving two independent referees in case of persisting disagreement. Another potential limitation is that the studies assessed in our rating may not cover all melanoma prediction studies. The basis for our set of studies were three systematic reviews that we supplemented with a literature update. Nevertheless, due to the eligibility criteria of the systematic reviews, we included in our assessment only studies reporting (i) solely the development and (ii) both the development and external validation of a melanoma risk prediction model. Thus, studies focusing exclusively on external validation or update of preexisting models, for which PROBAST is also designed, were not part of our investigation and our results hence do not allow conclusions regarding ROB in these study types.

5. Conclusions

In conclusion, the vast majority of studies on melanoma risk prediction models had a high ROB rating showing that the validity of published prediction models for incident cutaneous melanoma was poor. The selection of participants and the omission of appropriate validation efforts in the statistical analyses were frequent sources of bias. The resulting thin evidence base of high-quality studies made it impossible to answer the question of how to predict melanoma yet. However, the positive temporal trend in bias reduction inspires hope that this may change in the future.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Table S1: Decision rules for high and unclear PROBAST ratings.

Author Contributions: Conceptualization, O.G.; methodology, I.K., K.D., M.V.H., S.M., T.S. and O.G.; software, I.K.; validation, A.B.P.,W.U. and O.G.; formal analysis, I.K.; investigation, I.K., K.D., M.V.H., S.M., T.S. and O.G.; resources, O.G.; data curation, A.B.P. and I.K.; writing—original draft preparation, I.K.; writing—review and editing, C.B., K.D., M.V.H., S.M., A.B.P., T.S.,W.U. and O.G.; visualization, I.K.; supervision, O.G.; project administration, I.K.; funding acquisition, O.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest

 doi:10.20944/preprints202205.0091.v1

Appendix A

Table A 1: Study characteristics and PROBAST results per domain and overall of all included studies. Studies are ordered according to study type and year of publication. Within studies of the same study type and year of publication, the studies are sorted by the last name of the first author. (N=42)

Author	Study type	Publication year	Study design	ROB Rating				
				Participants	Predictors	Outcome	Analysis	Overall
English and Armstrong [30]	D	1988	Case-control	+	?	+	?	?
Garbe et al.[36]	D	1989	Case-control	-	?	+	-	-
MacKie et al.[44]	D	1989	Case-control	-	?	+	-	-
Augustsson et al.[21]	D	1991	Case-control	+	+	+	-	-
Marett et al.[46]	D	1992	Case-control	+	?	+	-	-
Garbe et al.[35]	D	1994	Case-control	-	?	+	-	-
Barbini et al.[23]	D	1998	Case-control	-	+	+	?	-
Landi et al.[43]	D	2001	Case-control	-	?	+	-	-
Harbauer et al.[40]	D	2003	Case-control	-	?	+	-	-
Dwyer et al.[29]	D	2004	Case-control	+	+	+	-	-
Fargnoli et al.[32]	D	2004	Case-control	-	?	+	-	-
Cho et al.[24]	D	2005	Cohort	-	-	+	+	-
Whiteman and Green[59]	D	2005	Published case-control studies	?	?	+	-	-
Fears et al.[33]	D	2006	Case-control	-	?	+	-	-
Goldberg et al.[37]	D	2007	Cohort	-	+	-	-	-
Mar et al.[45]	D	2011	Published meta-analysis and registry data	-	?	+	-	-
Nielsen et al.[47]	D	2011	Cohort	+	+	+	-	-
Quéreux et al.[51]	D	2011	Case-control	-	?	+	+	-
Williams et al.[60]	D	2011	Case-control	+	?	+	?	?
Guther et al.[39]	D	2012	Cohort	-	+	-	?	-
Smith et al.[53]	D	2012	Case-control	?	?	?	-	-
Bakos et al.[22]	D	2013	Case-control	-	?	+	-	-
Stefanaki et al.[55]	D	2013	Case-control	-	?	+	-	-
Nikolic et al.[48]	D	2014	Case-control	-	?	+	?	-

 doi:10.20944/preprints202205.0091.v1

Penn et al.[50]	D	2014	Case-control	+	?	+	?	?
Sneyd et al.[54]	D	2014	Case-control	+	?	+	+	-
Kypreou et al.[42]	D	2016	Case-control	-	+	+	+	-
Cho et al.[25]	D	2018	Cohort	+	+	-	+	-
Gu et al.[38]	D	2018	Case-control	-	-	+	?	-
Hübner et al.[41]	D	2018	Cohort study based on data form SCREEN project	-	+	+	-	-
Olsen et al.[49]	D	2018	Cohort study	+	+	+	+	+
Richter and Koshgof-taar[52]	D	2018	Cohort study based on EHR data	-	?	+	?	-
Tagliabue et al.[56]	D	2018	Case-control	-	-	+	-	-
Bakshi et al.[61]	D	2021	Cohort	+	+	+	-	-
Fontanillas et al.[62]	D	2021	Cohort	?	?	-	+	-
Fortes et al.[34]	D+V	2010	Case-control	-	?	+	+	-
Cust et al.[27]	D+V	2013	Case-control	+	?	+	+	?
Fang et al.[31]	D+V	2013	Multiple case-control studies	-	?	+	+	-
Davies et al.[28]	D+V	2015	Multiple case-control studies	-	+	+	+	-
Vuong et al.[58]	D+V	2016	Case-control	+	?	+	+	?
Cust et al.[26]	D+V	2018	Case-control	+	?	+	+	?
Vuong et al.[57]	D+V	2019	Case-control	+	?	+	+	?

Abbreviations: PROBAST = Prediction model Risk Of Bias ASsessment Tool; ROB = risk of bias; D = development studies; D+V = development and external validation studies; SCREEN = Skin Cancer Research to provide Evidence for Effectiveness of Screening in Northern Germany; EHR = Electronic Health Records; + indicates low ROB; - indicates high ROB; ? indicates unclear ROB

References

1. Matthews, N.H.; Li, W.Q.; Qureshi, A.A.; Weinstock, M.A.; Cho, E. Epidemiology of Melanoma. In *Cutaneous Melanoma: Etiology and Therapy*, Ward, W.H., Farma, J.M., Eds.; Brisbane (AU), 2017.
2. Garbe, C.; Keim, U.; Gandini, S.; Amaral, T.; Katalinic, A.; Hollezcek, B.; Martus, P.; Flatz, L.; Leiter, U.; Whiteman, D. Epidemiology of cutaneous melanoma and keratinocyte cancer in white populations 1943-2036. *Eur J Cancer* **2021**, *152*, 18-25, doi:10.1016/j.ejca.2021.04.029.
3. Leiter, U.; Keim, U.; Garbe, C. Epidemiology of Skin Cancer: Update 2019. *Adv Exp Med Biol* **2020**, *1268*, 123-139, doi:10.1007/978-3-030-46227-7_6.
4. Whiteman, D.C.; Green, A.C.; Olsen, C.M. The Growing Burden of Invasive Melanoma: Projections of Incidence Rates and Numbers of New Cases in Six Susceptible Populations through 2031. *J Invest Dermatol* **2016**, *136*, 1161-1171, doi:10.1016/j.jid.2016.01.035.
5. International Agency for Research on Cancer. Available online: https://gco.iarc.fr/today/online-analysis-table?v=2020&mode=population&mode_population=regions&population=900&populations=900&key=asr&sex=0&cancer=16&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=0&include_nmsc_other=1 (accessed on 03 March 2022).
6. Iannacone, M.R.; Green, A.C. Towards skin cancer prevention and early detection: evolution of skin cancer awareness campaigns in Australia. *Melanoma Manag* **2014**, *1*, 75-84, doi:10.2217/mmt.14.6.
7. Kaiser, I.; Pfahlberg, A.B.; Uter, W.; Heppt, M.V.; Veierod, M.B.; Gefeller, O. Risk Prediction Models for Melanoma: A Systematic Review on the Heterogeneity in Model Development and Validation. *Int J Environ Res Public Health* **2020**, *17*, doi:10.3390/ijerph17217919.
8. Usher-Smith, J.A.; Emery, J.; Kassianos, A.P.; Walter, F.M. Risk prediction models for melanoma: a systematic review. *Cancer Epidemiol Biomarkers Prev* **2014**, *23*, 1450-1463, doi:10.1158/1055-9965.EPI-14-0295.
9. Vuong, K.; McGeechan, K.; Armstrong, B.K.; Cust, A.E. Risk prediction models for incident primary cutaneous melanoma: a systematic review. *JAMA Dermatol* **2014**, *150*, 434-444, doi:10.1001/jamadermatol.2013.8890.
10. Siontis, G.C.; Tzoulaki, I.; Castaldi, P.J.; Ioannidis, J.P. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* **2015**, *68*, 25-34, doi:10.1016/j.jclinepi.2014.09.007.
11. Steyerberg, E.W. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*; Springer: 2009.
12. Mannan, H.R.; McNeil, J.J. Computer programs to estimate overoptimism in measures of discrimination for predicting the risk of cardiovascular diseases. *J Eval Clin Pract* **2013**, *19*, 358-362, doi:10.1111/j.1365-2753.2012.01834.x.
13. Wolff, R.F.; Moons, K.G.M.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S.; Groupdagger, P. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* **2019**, *170*, 51-58, doi:10.7326/M18-1376.
14. Moons, K.G.M.; Wolff, R.F.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* **2019**, *170*, W1-W33, doi:10.7326/M18-1377.
15. Kaiser, I.; Diehl, K.; Heppt, M.V.; Mathes, S.; Pfahlberg, A.B.; Steeb, T.; Uter, W.; Gefeller, O. Reporting Quality of Studies Developing and Validating Melanoma Prediction Models: An Assessment Based on the TRIPOD Statement. *Healthcare (Basel)* **2022**, *10*, doi:10.3390/healthcare10020238.
16. Wohlin, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In Proceedings of the Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, London, England, United Kingdom, 2014; pp. 1-10.

17. Leiner, D.J. SoSci Survey Version 3.2.21; Computer software. SoSci Survey GmbH: Munich, Germany, 2019; Available online: https://www.soscisurvey.de (accessed on 21 February 2022).
18. Mantel, N. Chi-Square Tests with One Degree of Freedom: Extensions of Mantel-Haenszel Procedure. *Journal of the American Statistical Association* **1963**, *58*, 690-700, doi:https://doi.org/2282717.
19. Mehta, C.R.a.P., Nitin R. A Network Algorithm for Performing Fisher's Exact Test in r x c Contingency Tables. *Journal of the American Statistical Association* **1983**, *78*, 427-434, doi:https://doi.org/2288652
20. R Development Core Team *R: A language and environment for statistical computing*, R Foundation for Statistical Computing: Vienna, Austria, 2021.
21. Augustsson, A. Melanocytic naevi, melanoma and sun exposure. *Acta Derm Venereol Suppl (Stockh)* **1991**, *166*, 1-34.
22. Bakos, L.; Mastroeni, S.; Bonamigo, R.R.; Melchi, F.; Pasquini, P.; Fortes, C. A melanoma risk score in a Brazilian population. *An Bras Dermatol* **2013**, *88*, 226-232, doi:10.1590/S0365-05962013000200007.
23. Barbini, P.; Cevenini, G.; Rubegni, P.; Massai, M.R.; Flori, M.L.; Carli, P.; Andreassi, L. Instrumental measurement of skin colour and skin type as risk factors for melanoma: a statistical classification procedure. *Melanoma Res* **1998**, *8*, 439-447, doi:10.1097/00008390-199810000-00009.
24. Cho, E.; Rosner, B.A.; Feskanich, D.; Colditz, G.A. Risk factors and individual probabilities of melanoma for whites. *J Clin Oncol* **2005**, *23*, 2669-2675, doi:10.1200/JCO.2005.11.108.
25. Cho, H.G.; Ransohoff, K.J.; Yang, L.; Hedlin, H.; Assimes, T.; Han, J.; Stefanick, M.; Tang, J.Y.; Sarin, K.Y. Melanoma risk prediction using a multilocus genetic risk score in the Women's Health Initiative cohort. *J Am Acad Dermatol* **2018**, *79*, 36-41 e10, doi:10.1016/j.jaad.2018.02.052.
26. Cust, A.E.; Drummond, M.; Kanetsky, P.A.; Australian Melanoma Family Study, I.; Leeds Case-Control Study, I.; Goldstein, A.M.; Barrett, J.H.; MacGregor, S.; Law, M.H.; Iles, M.M.; et al. Assessing the Incremental Contribution of Common Genomic Variants to Melanoma Risk Prediction in Two Population-Based Studies. *J Invest Dermatol* **2018**, *138*, 2617-2624, doi:10.1016/j.jid.2018.05.023.
27. Cust, A.E.; Goumas, C.; Vuong, K.; Davies, J.R.; Barrett, J.H.; Holland, E.A.; Schmid, H.; Agha-Hamilton, C.; Armstrong, B.K.; Kefford, R.F.; et al. MC1R genotype as a predictor of early-onset melanoma, compared with self-reported and physician-measured traditional risk factors: an Australian case-control-family study. *BMC Cancer* **2013**, *13*, 406, doi:10.1186/1471-2407-13-406.
28. Davies, J.R.; Chang, Y.M.; Bishop, D.T.; Armstrong, B.K.; Bataille, V.; Bergman, W.; Berwick, M.; Bracci, P.M.; Elwood, J.M.; Ernstoff, M.S.; et al. Development and validation of a melanoma risk score based on pooled data from 16 case-control studies. *Cancer Epidemiol Biomarkers Prev* **2015**, *24*, 817-824, doi:10.1158/1055-9965.EPI-14-1062.
29. Dwyer, T.; Stankovich, J.M.; Blizzard, L.; FitzGerald, L.M.; Dickinson, J.L.; Reilly, A.; Williamson, J.; Ashbolt, R.; Berwick, M.; Sale, M.M. 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, doi:10.1093/aje/kwh120.
30. English, D.R.; Armstrong, B.K. 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, doi:10.1136/bmj.296.6632.1285.
31. Fang, S.; Han, J.; Zhang, M.; Wang, L.E.; Wei, Q.; Amos, C.I.; Lee, J.E. Joint effect of multiple common SNPs predicts melanoma susceptibility. *PLoS One* **2013**, *8*, e85642, doi:10.1371/journal.pone.0085642.
32. Fargnoli, M.C.; Piccolo, D.; Altobelli, E.; Formicone, F.; Chimenti, S.; Peris, K. Constitutional and environmental risk factors for cutaneous melanoma in an Italian population. A case-control study. *Melanoma Res* **2004**, *14*, 151-157, doi:10.1097/00008390-200404000-00013.

33. Fears, T.R.; Guerry, D.t.; Pfeiffer, R.M.; Sagebiel, R.W.; Elder, D.E.; Halpern, A.; Holly, E.A.; Hartge, P.; Tucker, M.A. Identifying individuals at high risk of melanoma: a practical predictor of absolute risk. *J Clin Oncol* **2006**, *24*, 3590-3596, doi:10.1200/JCO.2005.04.1277.
34. Fortes, C.; Mastroeni, S.; Bakos, L.; Antonelli, G.; Alessandroni, L.; Pilla, M.A.; Alotto, M.; Zappala, A.; Manoorannparampill, T.; Bonamigo, R.; et al. Identifying individuals at high risk of melanoma: a simple tool. *Eur J Cancer Prev* **2010**, *19*, 393-400, doi:10.1097/CEJ.0b013e32833b492f.
35. Garbe, C.; Buttner, P.; Weiss, J.; Soyer, H.P.; Stocker, U.; Kruger, S.; Roser, M.; Weckbecker, J.; Panizzon, R.; Bahmer, F.; 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*, 695-699, doi:10.1111/1523-1747.ep12374280.
36. Garbe, C.; Kruger, S.; Stadler, R.; Guggenmoos-Holzmann, I.; Orfanos, C.E. Markers and relative risk in a German population for developing malignant melanoma. *Int J Dermatol* **1989**, *28*, 517-523, doi:10.1111/j.1365-4362.1989.tb04604.x.
37. Goldberg, M.S.; Doucette, J.T.; Lim, H.W.; Spencer, J.; Carucci, J.A.; Rigel, D.S. 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, doi:10.1016/j.jaad.2007.02.010.
38. Gu, F.; Chen, T.H.; Pfeiffer, R.M.; Fargnoli, M.C.; Calista, D.; Ghiorzo, P.; Peris, K.; Puig, S.; Menin, C.; De Nicolo, A.; et al. Combining common genetic variants and non-genetic risk factors to predict risk of cutaneous melanoma. *Hum Mol Genet* **2018**, *27*, 4145-4156, doi:10.1093/hmg/ddy282.
39. 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 Dermatol Venereol* **2012**, *26*, 86-94, doi:10.1111/j.1468-3083.2011.04014.x.
40. 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, doi:10.1097/00008390-200310000-00013.
41. Hubner, J.; Waldmann, A.; Eisemann, N.; Noftz, M.; Geller, A.C.; Weinstock, M.A.; Volkmer, B.; Greinert, R.; Breitbart, E.W.; Katalinic, A. Association between risk factors and detection of cutaneous melanoma in the setting of a population-based skin cancer screening. *Eur J Cancer Prev* **2018**, *27*, 563-569, doi:10.1097/CEJ.0000000000000392.
42. Kypreou, K.P.; Stefanaki, I.; Antonopoulou, K.; Karagianni, F.; Ntritsos, G.; Zaras, A.; Nikolaou, V.; Kalfa, I.; Chasapi, V.; Polydorou, D.; et al. Prediction of Melanoma Risk in a Southern European Population Based on a Weighted Genetic Risk Score. *J Invest Dermatol* **2016**, *136*, 690-695, doi:10.1016/j.jid.2015.12.007.
43. Landi, M.T.; Baccarelli, A.; Calista, D.; Pesatori, A.; Fears, T.; Tucker, M.A.; Landi, G. Combined risk factors for melanoma in a Mediterranean population. *Br J Cancer* **2001**, *85*, 1304-1310, doi:10.1054/bjoc.2001.2029.
44. MacKie, R.M.; Freudenberger, T.; Aitchison, T.C. Personal risk-factor chart for cutaneous melanoma. *Lancet* **1989**, *2*, 487-490, doi:10.1016/s0140-6736(89)92097-7.
45. Mar, V.; Wolfe, R.; Kelly, J.W. Predicting melanoma risk for the Australian population. *Australas J Dermatol* **2011**, *52*, 109-116, doi:10.1111/j.1440-0960.2010.00727.x.
46. Marrett, L.D.; King, W.D.; Walter, S.D.; From, L. Use of Host Factors to Identify People at High-Risk for Cutaneous Malignant-Melanoma. *Can Med Assoc J* **1992**, *147*, 445-452.
47. 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. *Int J Cancer* **2012**, *131*, 706-715, doi:10.1002/ijc.26408.
48. Nikolic, J.; Loncar-Turukalo, T.; Sladojevic, S.; Marinkovic, M.; Janjic, Z. Melanoma risk prediction models. *Vojnosanit Pregl* **2014**, *71*, 757-766, doi:10.2298/vsp130722045n.

49. Olsen, C.M.; Pandeya, N.; Thompson, B.S.; Dusingize, J.C.; Webb, P.M.; Green, A.C.; Neale, R.E.; Whiteman, D.C.; Study, Q.S. Risk Stratification for Melanoma: Models Derived and Validated in a Purpose-Designed Prospective Cohort. *J Natl Cancer Inst* **2018**, *110*, 1075-1083, doi:10.1093/jnci/djy023.
50. Penn, L.A.; Qian, M.; Zhang, E.; Ng, E.; Shao, Y.; Berwick, M.; Lazovich, D.; Polsky, D. Development of a melanoma risk prediction model incorporating MC1R genotype and indoor tanning exposure: impact of mole phenotype on model performance. *PLoS One* **2014**, *9*, e101507, doi:10.1371/journal.pone.0101507.
51. Quereux, G.; Moyse, D.; Lequeux, Y.; Jumbou, O.; Brocard, A.; Antonioli, D.; Dreno, B.; Nguyen, J.M. Development of an individual score for melanoma risk. *Eur J Cancer Prev* **2011**, *20*, 217-224, doi:10.1097/CEJ.0b013e32834474ae.
52. Richter, A.; Khoshgoftaar, T. Melanoma Risk Prediction with Structured Electronic Health Records. In Proceedings of the ACM-BCB'18: 9th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 2018.
53. Smith, L.A.; Qian, M.; Ng, E.; Shao, Y.Z.; Berwick, M.; Lazovich, D.; Polsky, D. Development of a melanoma risk prediction model incorporating MC1R genotype and indoor tanning exposure. *J Clin Oncol* **2012**, *30*, 8574-8574, doi:10.1200/jco.2012.30.15_suppl.8574.
54. Sneyd, M.J.; Cameron, C.; Cox, B. Individual risk of cutaneous melanoma in New Zealand: developing a clinical prediction aid. *BMC Cancer* **2014**, *14*, 359, doi:10.1186/1471-2407-14-359.
55. Stefanaki, I.; Panagiotou, O.A.; Kodela, E.; Gogas, H.; Kypreou, K.P.; Chatzinasiou, F.; Nikolaou, V.; Plaka, M.; Kalfa, I.; Antoniou, C.; et al. Replication and predictive value of SNPs associated with melanoma and pigmentation traits in a Southern European case-control study. *PLoS One* **2013**, *8*, e55712, doi:10.1371/journal.pone.0055712.
56. Tagliabue, E.; Gandini, S.; Bellocco, R.; Maisonneuve, P.; Newton-Bishop, J.; Polsky, D.; Lazovich, D.; Kanetsky, P.A.; Ghiorzo, P.; Gruis, N.A.; et al. MC1R variants as melanoma risk factors independent of at-risk phenotypic characteristics: a pooled analysis from the M-SKIP project. *Cancer Manag Res* **2018**, *10*, 1143-1154, doi:10.2147/CMAR.S155283.
57. Vuong, K.; Armstrong, B.K.; Drummond, M.; Hopper, J.L.; Barrett, J.H.; Davies, J.R.; Bishop, D.T.; Newton-Bishop, J.; Aitken, J.F.; Giles, G.G.; et al. Development and external validation study of a melanoma risk prediction model incorporating clinically assessed naevi and solar lentigines. *Br J Dermatol* **2020**, *182*, 1262-1268, doi:10.1111/bjd.18411.
58. Vuong, K.; Armstrong, B.K.; Weiderpass, E.; Lund, E.; Adami, H.O.; Veierod, M.B.; Barrett, J.H.; Davies, J.R.; Bishop, D.T.; Whiteman, D.C.; et al. Development and External Validation of a Melanoma Risk Prediction Model Based on Self-assessed Risk Factors. *JAMA Dermatol* **2016**, *152*, 889-896, doi:10.1001/jamadermatol.2016.0939.
59. Whiteman, D.C.; Green, A.C. A risk prediction tool for melanoma? *Cancer Epidemiol Biomarkers Prev* **2005**, *14*, 761-763, doi:10.1158/1055-9965.EPI-14-4-ED.
60. Williams, L.H.; Shors, A.R.; Barlow, W.E.; Solomon, C.; White, E. Identifying Persons at Highest Risk of Melanoma Using Self-Assessed Risk Factors. *J Clin Exp Dermatol Res* **2011**, *2*, doi:10.4172/2155-9554.1000129.
61. Bakshi, A.; Yan, M.; Riaz, M.; Polekhina, G.; Orchard, S.G.; Tiller, J.; Wolfe, R.; Joshi, A.; Cao, Y.; McInerney-Leo, A.M.; et al. Genomic Risk Score for Melanoma in a Prospective Study of Older Individuals. *J Natl Cancer Inst* **2021**, *113*, 1379-1385, doi:10.1093/jnci/djab076.
62. Fontanillas, P.; Alipanahi, B.; Furlotte, N.A.; Johnson, M.; Wilson, C.H.; 23andMe Research Team; Pitts, S.J.; Gentleman, R.; Auton, A. Disease risk scores for skin cancers. *Nat Commun* **2021**, *12*, 160, doi:10.1038/s41467-020-20246-5.
63. Olsen, C.M.; Green, A.C.; Neale, R.E.; Webb, P.M.; Cicero, R.A.; Jackman, L.M.; O'Brien, S.M.; Perry, S.L.; Ranieri, B.A.; Whiteman, D.C.; et al. Cohort profile: the QSkin Sun and Health Study. *Int J Epidemiol* **2012**, *41*, 929-929i, doi:10.1093/ije/dys107.
64. Cust, A.E.; Schmid, H.; Maskiell, J.A.; Jetann, J.; Ferguson, M.; Holland, E.A.; Agha-Hamilton, C.; Jenkins, M.A.; Kelly, J.; Kefford, R.F.; et al. Population-based, Case-Control-Family Design to Investigate Genetic and Environmental Influences on Melanoma Risk Australian Melanoma Family Study. *Am J Epidemiol* **2009**, *170*, 1541-1554, doi:10.1093/aje/kwp307.

65. Newton-Bishop, J.A.; Chang, Y.M.; Iles, M.M.; Taylor, J.C.; Bakker, B.; Chan, M.; Leake, S.; Karpavicius, B.; Haynes, S.; Fitzgibbon, E.; et al. Melanocytic Nevi, Nevus Genes, and Melanoma Risk in a Large Case-Control Study in the United Kingdom. *Cancer Epidem Biomar* **2010**, *19*, 2043-2054, doi:10.1158/1055-9965.Epi-10-0233.
66. Cockburn, M.; Hamilton, A.; Mack, T. Recall bias in self-reported melanoma risk factors. *Am J Epidemiol* **2001**, *153*, 1021-1026, doi:10.1093/aje/153.10.1021.
67. de Vries, E.; Boniol, M.; Severi, G.; Eggermont, A.M.; Autier, P.; Bataille, V.; Dore, J.F.; Coebergh, J.W. Public awareness about risk factors could pose problems for case-control studies: the example of sunbed use and cutaneous melanoma. *Eur J Cancer* **2005**, *41*, 2150-2154, doi:10.1016/j.ejca.2005.04.042.
68. Han, J.; Colditz, G.A.; Hunter, D.J. Risk factors for skin cancers: a nested case-control study within the Nurses' Health Study. *Int J Epidemiol* **2006**, *35*, 1514-1521, doi:10.1093/ije/dyl197.
69. Parr, C.L.; Hjartaker, A.; Laake, P.; Lund, E.; Veierod, M.B. Recall bias in melanoma risk factors and measurement error effects: a nested case-control study within the Norwegian Women and Cancer Study. *Am J Epidemiol* **2009**, *169*, 257-266, doi:10.1093/aje/kwn363.
70. Veierod, M.B.; Parr, C.L.; Lund, E.; Hjartaker, A. Reproducibility of self-reported melanoma risk factors in a large cohort study of Norwegian women. *Melanoma Res* **2008**, *18*, 1-9, doi:10.1097/CMR.0b013e3282f120d2.
71. Walter, S.D.; Marrett, L.D.; Shannon, H.S.; From, L.; Hertzman, C. The association of cutaneous malignant melanoma and fluorescent light exposure. *Am J Epidemiol* **1992**, *135*, 749-762, doi:10.1093/oxfordjournals.aje.a116361.
72. Weinstock, M.A.; Colditz, G.A.; Willett, W.C.; Stampfer, M.J.; Rosner, B.; Speizer, F.E. Recall (report) bias and reliability in the retrospective assessment of melanoma risk. *Am J Epidemiol* **1991**, *133*, 240-245, doi:10.1093/oxfordjournals.aje.a115868.
73. Gefeller, O. Invited commentary: Recall bias in melanoma -- much ado about almost nothing? *Am J Epidemiol* **2009**, *169*, 267-270; discussion 271-262, doi:10.1093/aje/kwn362.
74. Pfahlberg, A.B.; Gefeller, O. Errors in assessing risk factors for melanoma: lack of reproducibility is the minor problem. *Melanoma Res* **2008**, *18*, 300-301, doi:10.1097/CMR.0b013e328308da8e.
75. Schulz, K.F.; Grimes, D.A. Case-control studies: research in reverse. *Lancet* **2002**, *359*, 431-434, doi:10.1016/S0140-6736(02)07605-5.
76. Heinemann, L.A.J.; Lewis, M.A.; Assmann, A.; Thiel, C. Case-control studies on venous thromboembolism: bias due to design? A methodological study on venous thromboembolism and steroid hormone use. *Contraception* **2002**, *65*, 207-214, doi:10.1016/S0010-7824(01)00309-2.
77. Patten, S.B. Selection bias in studies of major depression using clinical subjects. *J Clin Epidemiol* **2000**, *53*, 351-357, doi:Doi 10.1016/S0895-4356(99)00215-2.
78. Stavraky, K.M.; Clarke, E.A. Hospital or Population-Controls - an Unanswered Question. *J Chron Dis* **1983**, *36*, 301-307, doi:Doi 10.1016/0021-9681(83)90113-3.
79. Dantas, O.M.S.; Ximenes, R.A.A.; de Albuquerque, M.D.P.M.; Montarroyos, U.R.; de Souza, W.V.; Varejao, P.; Rodrigues, L.C. Selection bias: neighbourhood controls and controls selected from those presenting to a Health Unit in a case control study of efficacy of BCG revaccination. *BMC Med Res Methodol* **2007**, *7*, doi:10.1186/1471-2288-7-11.
80. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* **2015**, *350*, g7594, doi:10.1136/bmj.g7594.
81. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *J Clin Epidemiol* **2021**, *134*, 178-189, doi:10.1016/j.jclinepi.2021.03.001.

82. Liu, H.; Zhou, X.; Yu, G.; Sun, X. The effects of the PRISMA statement to improve the conduct and reporting of systematic reviews and meta-analyses of nursing interventions for patients with heart failure. *Int J Nurs Pract* **2019**, *25*, e12729, doi:10.1111/ijn.12729.

83. Higgins, J.P.; Altman, D.G.; Gotzsche, P.C.; Juni, P.; Moher, D.; Oxman, A.D.; Savovic, J.; Schulz, K.F.; Weeks, L.; Sterne, J.A.; et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* **2011**, *343*, d5928, doi:10.1136/bmj.d5928.

84. Sterne, J.A.C.; Savovic, J.; Page, M.J.; Elbers, R.G.; Blencowe, N.S.; Boutron, I.; Cates, C.J.; Cheng, H.Y.; Corbett, M.S.; Eldridge, S.M.; et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* **2019**, *366*, l4898, doi:10.1136/bmj.l4898.

85. Sassano, M.; Mariani, M.; Quaranta, G.; Pastorino, R.; Boccia, S. Polygenic risk prediction models for colorectal cancer: a systematic review. *BMC Cancer* **2022**, *22*, 65, doi:10.1186/s12885-021-09143-2.

86. Su, N.; Lagerweij, M.D.; van der Heijden, G. Assessment of predictive performance of caries risk assessment models based on a systematic review and meta-analysis. *J Dent* **2021**, *110*, 103664, doi:10.1016/j.jdent.2021.103664.

87. de Jong, Y.; Ramspek, C.L.; Zoccali, C.; Jager, K.J.; Dekker, F.W.; van Diepen, M. 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*, 939-947, doi:10.1111/nep.13913.

88. Austin, H.; Hill, H.A.; Flanders, W.D.; Greenberg, R.S. Limitations in the application of case-control methodology. *Epidemiol Rev* **1994**, *16*, 65-76, doi:10.1093/oxfordjournals.epirev.a036146.