International Journal of
Environmental Research and Public Health

MDPI

Review

Risk Prediction Models for Melanoma: A Systematic Review on the Heterogeneity in Model Development and Validation

Isabelle Kaiser [1], Annette B. Pfahlberg [1], Wolfgang Uter [1], Markus V. Heppt [2], Marit B. Veierød [3] and Olaf Gefeller [1,*]

1 Department of Medical Informatics, Biometry and Epidemiology, Friedrich Alexander University of Erlangen-Nuremberg, 91054 Erlangen, Germany; isabelle.kaiser@fau.de (I.K.); annette.pfahlberg@fau.de (A.B.P.); wolfgang.uter@fau.de (W.U.)
2 Department of Dermatology, University Hospital Erlangen, 91054 Erlangen, Germany; markus.heppt@uk-erlangen.de
3 Oslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, 0317 Oslo, Norway; m.b.veierod@medisin.uio.no
* Correspondence: olaf.gefeller@fau.de

Received: 14 September 2020; Accepted: 26 October 2020; Published: 28 October 2020

Abstract: The rising incidence of cutaneous melanoma over the past few decades has prompted substantial efforts to develop risk prediction models identifying people at high risk of developing melanoma to facilitate targeted screening programs. We review these models, regarding study characteristics, differences in risk factor selection and assessment, evaluation, and validation methods. Our systematic literature search revealed 40 studies comprising 46 different risk prediction models eligible for the review. Altogether, 35 different risk factors were part of the models with nevi being the most common one (n = 35, 78%); little consistency in other risk factors was observed. Results of an internal validation were reported for less than half of the studies (n = 18, 45%), and only 6 performed external validation. In terms of model performance, 29 studies assessed the discriminative ability of their models; other performance measures, e.g., regarding calibration or clinical usefulness, were rarely reported. Due to the substantial heterogeneity in risk factor selection and assessment as well as methodologic aspects of model development, direct comparisons between models are hardly possible. Uniform methodologic standards for the development and validation of risk prediction models for melanoma and reporting standards for the accompanying publications are necessary and need to be obligatory for that reason.

Keywords: melanoma; risk prediction; statistical models; validation

1. Introduction

Cutaneous melanoma is one of the most lethal forms of skin cancer and accounts for the majority of skin cancer deaths [1]. Over the past few decades, the incidence of melanoma has risen dramatically worldwide, especially in regions with fair-skinned populations [1–3]. While in many European populations, like the UK and the Netherlands, incidence rates increased with estimated annual percentage changes of 4 to 6% event in recent decades, the annual increase in incidence in Australia and New Zealand seems to have leveled off since 1995 [4]. Nonetheless, Australia and New Zealand have the highest incidence rates worldwide, followed by Western Europe, Northern Europe, and the USA [1,3,4]. In 2018, melanoma was the fourth most common cancer in Australia regarding incidence [5]. In Europe (except Southern Europe) and the USA, it ranks fifth and sixth [5].

Melanoma survival is strongly related to tumor thickness at diagnosis and whether the tumor has already metastasized or not [6]. Therefore, early diagnosis is important for the successful treatment of the disease and to keep mortality at a low level [1,2,7].

Melanoma screening programs are essential for early diagnosis [2]. Although population-based screening efforts have the potential to decrease the mortality rate of melanoma, as shown by the SCREEN (Skin Cancer Research to Provide Evidence for Effectiveness of Screening in Northern Germany) project [7,8], screening the whole population is cost-intensive and its efficacy unproven [9,10]. Thus, targeted screening of individuals at higher risk of developing melanoma is regarded as the more effective prevention strategy [11,12]. Targeted screening needs proper identification of the subgroup with a higher melanoma risk. To this end, so called risk prediction models for melanoma have been developed [13]. Based on a combination of risk factors, they assess an individual's risk of developing melanoma. Tailored risk estimates provided by prediction models have the additional advantage that individuals may interpret this information as more personally relevant for them and may feel encouraged to participate in melanoma screening programs or even favorably change their behavior in order to reduce melanoma risk [14,15].

Due to the growing general interest in cancer risk prediction and the increasing incidence rates of melanoma, a high number of melanoma risk prediction models have been developed over the last few decades. However, the more models are being developed, the more difficult it is for practitioners to judge which model is the most useful [15]. Therefore, details on the performance and reliability of the models, as well as direct comparisons of multiple models, are necessary in order to provide guidance as to which model to use in clinical practice. For melanoma, however, quantitative comparisons of risk prediction models are hitherto limited.

The aim of this systematic review is to give an overview of published risk prediction models for melanoma while highlighting the heterogeneity in definition and assessment of predictors incorporated in these models. Additionally, we describe systematically how these models have been validated and which methods and measures to evaluate their performance have been employed.

2. Materials and Methods

2.1. Eligibility Criteria and Search Strategy

For our analysis, we searched for studies describing prediction models quantifying the risk of melanoma and identifying people at high risk of developing melanoma, respectively. Only studies using a multivariable prediction model were eligible for this review. The models should provide either absolute risks or risk scores, or report the mutually adjusted relative risks of individual risk factors and risk factor combinations in multifactorial statistical models. As we focus on the prediction of new cases instead of the recurrence of melanoma, we concentrated on models for primary cutaneous melanoma (both invasive and in situ). Furthermore, only studies using a well-defined statistical method for the development of their models and risk predictions derived from these models, respectively, were eligible for this review, while models developed primarily based on expert opinions or consensus meetings were excluded.

We started our search with the systematic reviews of Vuong et al. [15] and Usher-Smith et al. [13], which are the only systematic reviews regarding melanoma risk prediction. They comprise a total of 26 different studies [16–41]. Both systematic reviews were published in 2014 and used similar eligibility criteria. They only included studies developing multivariable prediction models for incident primary cutaneous melanoma providing a measure of relative or absolute risk for members of the general population. Further eligibility criteria of Usher-Smith et al. [13] were that the studies were published as primary research articles and derived the final risk prediction model from their own study data. In contrast, Vuong et al. [15] made no specific restrictions on publication type, study type, and methodology. We adopted the less restrictive eligibility criteria of Vuong et al. [15], but had to exclude two of the 26 studies [38,41], both appearing only in Usher-Smith et al. [13], from further

consideration. One study publication [41] is only available in Russian language and could thus not be assessed by us. The other study [38] focused on the identification of risk factors rather than on the prediction of melanoma risk and the identification of people with a high melanoma risk, respectively. The study did not report a prediction model containing mutually adjusted relative risks and thus did not fulfill the eligibility criteria. We performed forward snowballing (searching within the citations to the paper being examined [42]) on these two systematic reviews using Google Scholar and Web of Science. Due to its reported high precision, forward snowballing is a useful approach to support the update of systematic literature reviews [42–44], which is why we used this method to find more recent melanoma risk prediction studies. Additionally, we performed an electronic literature search in PubMed from May 1, 2013 (as this is the date the literature search of Vuong et al. [15] ended) up to January 31, 2020. We used combinations of the search terms "risk/risk assessment/probability", "prediction/model/score" and "melanoma/skin cancer" (see Supplementary File S1 for complete search strategy).

2.2. Data Extraction

In case of studies developing several models by starting with a base model and adding more risk factors step-by-step, only the full model has been included. Regarding studies with several models but without a differentiation between a base and a full model, the model with the best performance was included. If there was no model with significantly better performance, all models were selected for analysis. For studies with separate models for men and women, or for self-assessment and physician assessment, both models were included separately.

The characteristics of study design, study sample, analytic model, predictive factors included in final models, and outcome measures were abstracted, as well as evaluation and performance measures. Additionally, spatio-temporal information on the location and time of the studies were extracted.

2.3. Data Processing

The studies and their prediction models, respectively, were analyzed regarding the following aspects: (1) the spatio-temporal information; (2) the heterogeneity of risk factors in general; (3) the disparities in defining and ascertaining individual risk factors; (4) the validation methods, and (5) the evaluation of model performance.

To provide aggregated information on the geographical location, we allocated the studies to the countries and continents where their data sets originate. Additionally, we performed a temporal synthesis by dividing the entire period of time, in which the studies were published, into eight intervals of four years each.

To report the overall heterogeneity of risk factors, classification of model variables into meaningful groups was performed in order to keep the amount of different risk factors manageable. Variables which have the same meaning but different names (e.g., dysplastic nevi and atypical nevi) were aggregated. In addition, variables that belong to the same topic (e.g., sunburns in childhood, lifetime sunburns and sunburns without further specification) were also combined. One study used a pigmentation score as risk factor that was calculated reproducibly from the variables hair color, eye color, tanning ability and skin color [45]. We therefore applied the four individual factors in our analysis instead of the score. Concerning the variable red hair (RH)-phenotype used in [46], which is a combination of the risk factors hair color, freckles and Fitzpatrick skin type, we proceeded similarly by including the three phenotypic characteristics as separate variables. A full list of all adaptions made is presented in the Supplementary Table S1.

To illustrate that the assessment of the same risk factor was not uniform in all studies, the variable nevi was analyzed in detail regarding differences in the size of the nevi that were counted, the body site on which they were counted, the examiner who was counting and the measurement level of the variable. The variable nevi was chosen as it is the most common risk factor in melanoma risk prediction models. Another example is the risk factor sunburns, which we also examined regarding its diverse definitions and measurement levels.

Validation methods comprise internal and external validation. While internal validation relates to the reproducibility of the model and should be performed during model development, external validation refers to the generalizability of the model to other populations [47]. Possible internal validation methods are split sample validation, bootstrap resampling and cross-validation.

Evaluation techniques involve measures of model performance like discrimination and calibration, which are traditional approaches to determine the performance of prediction models. In addition, other more recently suggested performance measures related to the clinical usefulness of a prediction model and reclassification like decision curves and the net reclassification index (NRI) were also considered [48]. Furthermore, we analyzed the difference between older and newer studies regarding both validation and evaluation methods. Therefore, we divided the included studies based on the median year of publication into two equal groups ("studies published up to 2011" and "studies published after 2011").

The results were summarized using descriptive statistics; frequencies and percentages were gathered and displayed in tabular form. Percentages were relative to the total number of studies or risk prediction models. It must be noted that the total number of studies differs from the total number of risk prediction models, as some studies developed multiple models. In case of subgroup analyses, e.g., when only models including the risk factor nevi are considered, percentages relate to the number of models in this subgroup.

3. Results

3.1. Study Selection

Altogether, 24 of the 26 studies [16–37,39,40] from the two systematic reviews [13,15] were included in our analysis. Eight further studies were identified via forward snowballing [46,49–55], while eight studies were found in PubMed [45,56–62]. Therefore, we included 40 studies in our analysis (Figure 1).

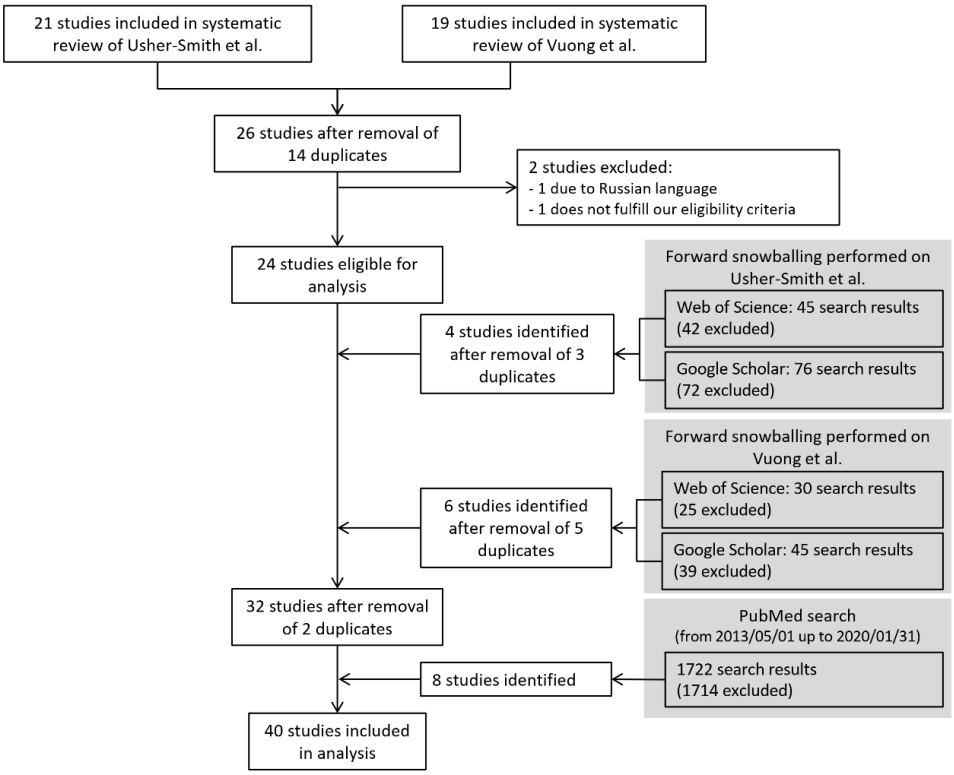

Figure 1. Flow diagram for the identification of studies developing risk prediction models for melanoma.

Thirteen of these studies developed multiple models, up to six models in one study [56]. Thus, the total number of models described in the 40 studies amounted to 66. By only including full models and models that have the best performance, when possible, we reduced the number of risk prediction models to 46.

3.2. Study Characteristics

Spatio-temporal information on the location and time of the studies were aggregated and are shown in Figures 2 and 3. Figure 2 displays the distribution of studies according to their continent and country of origin. The majority of studies (n = 36) originated from countries under the top 20 with the highest melanoma incidence rates, which include the United States (n = 9), Australia (n = 8), and Germany (n = 4) [63]. Four studies used data sets from several countries with high melanoma incidences for the development of their model.

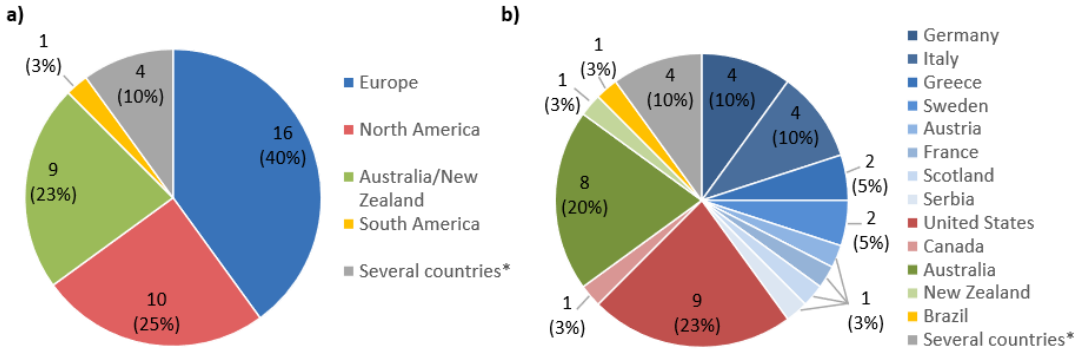

Figure 2. Geographical location of studies. (**a**) Distribution of studies according to the continents of their origin. (**b**) Distribution of studies according to their country of origin. (n = 40 studies). * Four studies used data sets from multiple countries with high melanoma incidences for the development of their risk prediction model.

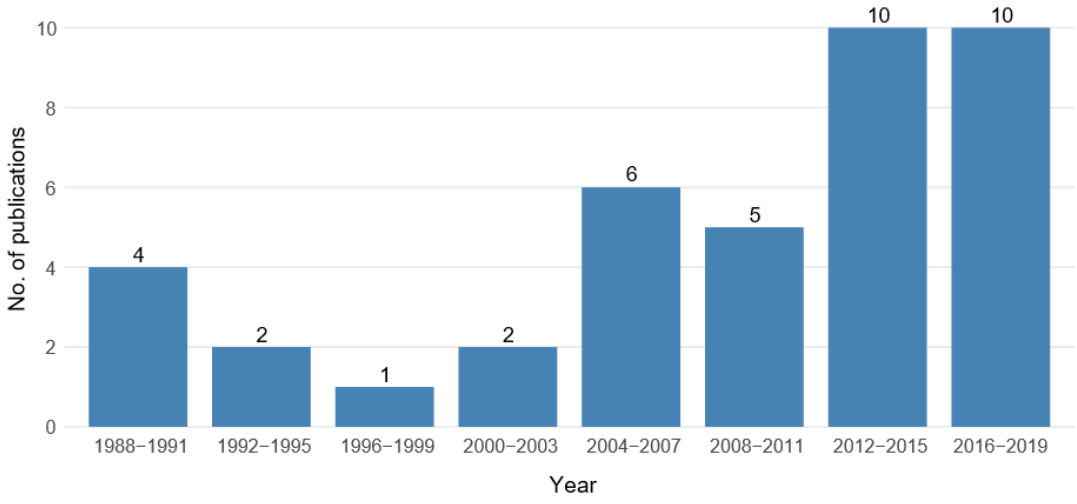

Figure 3. Temporal distribution of the reviewed studies showing the number of publications in eight time intervals of four years each. (n = 40 studies).

The temporal distribution of the reviewed studies is shown in Figure 3. The most recent time intervals, 2012–2015 and 2016–2019, show the highest numbers of publications (n = 10 each). In fact, half of the studies were published in those eight years, while the remaining 20 studies are spread over the larger time period from 1988 to 2011.

The key information extracted from each study is summarized in Table A1. Study designs used were mainly case-control (n = 30) and cohort (n = 8). Two studies used published material

from meta-analyses to obtain risk estimates. Most of the studies used logistic regression for model development (n = 29), but also the Gail method and Cox regression were applied several times (n = 5 and n = 3 studies, respectively). The Gail method was originally developed to assess the risk of breast cancer and combines risk estimates with incidence and mortality rates [15]. Two studies used the machine learning techniques random forest and decision trees to build their risk prediction models. In three studies, multiple approaches were employed to develop different models. Regarding the reported risk measure, only eight studies calculate the absolute risk for an individual to develop melanoma. A total of 12 studies calculated a risk score, of which 10 defined a cut-off point for being at high risk. Three further studies used relative risks with definitions of high risk. All other studies calculated the relative risk or odds ratio of single predictors or factor combinations without giving a definition of when a person is considered as being at high risk.

The average number of variables incorporated in a model was six (range 1–16). Cho et al. [58] developed their risk prediction model based on only one variable, a genetic risk score (GRS) which comprised aggregated genetic information from 21 single nucleotide polymorphisms (SNPs). All other studies described models including multiple independent variables. One study [55] used data of electronic health records (EHR) and therefore included a large number of variables in the risk prediction model; however, the exact number was not reported. As most of the risk factors were not specified in the publication, this study was excluded in the analysis of predictors included in the models.

3.3. Risk Factors Included in the Prediction Models

After grouping all risk factors in categories as described in Supplementary Table S1, we obtained 35 different predictors. Besides phenotypic risk factors like nevi, eye color and Fitzpatrick skin type, genetic and demographic risk factors were also used, as well as risk factors related to sun exposure and pigmented lesions (Table 1). The most common predictor is nevi, which has been used in 78% of the risk prediction models, followed by hair color (58%). Polygenic risk scores (PRS) were used in five models. In total, 15 of the 35 risk factors only occur in one or a maximum of two models. Almost all studies (except [58] and [26]) included at least one phenotypic predictor, whereas only one third of the models included genetic risk factors.

Table 1. Absolute (n) and relative frequencies (%) of predictive factors included in the risk prediction models for melanoma (n = 45 models *).

Risk Factors	*n*	%
Phenotypic factors		
Nevi	35	77.8
Hair color	26	57.8
Fitzpatrick	17	37.8
Freckles	16	35.6
Skin color	15	33.3
Eye color	14	31.1
Tanning ability	10	22.2
Genetic factors		
MC1R genotype	7	15.6
Polygenic risk score	5	11.1
SNPs	1	2.2
Demographic factors		
Age	16	35.6
Sex	15	33.3
Family history of melanoma	13	28.9
Residence	3	6.7
Level of education	1	2.2
Country of birth	1	2.2
Health insurance	1	2.2
Ethnicity	1	2.2
1st degree relative with large or unusual moles	1	2.2

Table 1. *Cont.*

Risk Factors	*n*	%
Sun exposure		
Sunburns	13	28.9
Sunbed sessions	7	15.6
Sun exposure	7	15.6
Occupational sun exposure	2	4.4
Use of sunscreen	2	4.4
Skin lesions		
Non-melanoma skin cancer	10	22.2
Atypical nevi	10	22.2
Sun damage	8	17.8
Melanoma history	5	11.1
Congenital nevi	2	4.4
Previous skin lesions treated destructively	2	4.4
Suspicious melanocytic lesions	1	2.2
Changing moles	1	2.2
Other risk factors		
Skin checks	2	4.4
Hormonal contraceptive therapy	1	2.2
Age on arrival in Australia	1	2.2

Abbreviations: MC1R = melanocortin 1 receptor, SNP = Single Nucleotide Polymorphism. * Study of Richter et al. [55] excluded due to limited reporting of predictors.

Figure 4 displays the frequencies of risk factor combinations as a heatmap. This figure only contains those 20 risk factors that appear in more than two of the risk prediction models. A plot with all 35 risk factors can be found in the Supplementary Figure S1. The most common predictor combinations are nevi and hair color (n = 21, 47%), followed by nevi and freckles (n = 15, 33%) and nevi and Fitzpatrick skin type (n = 14, 31%).

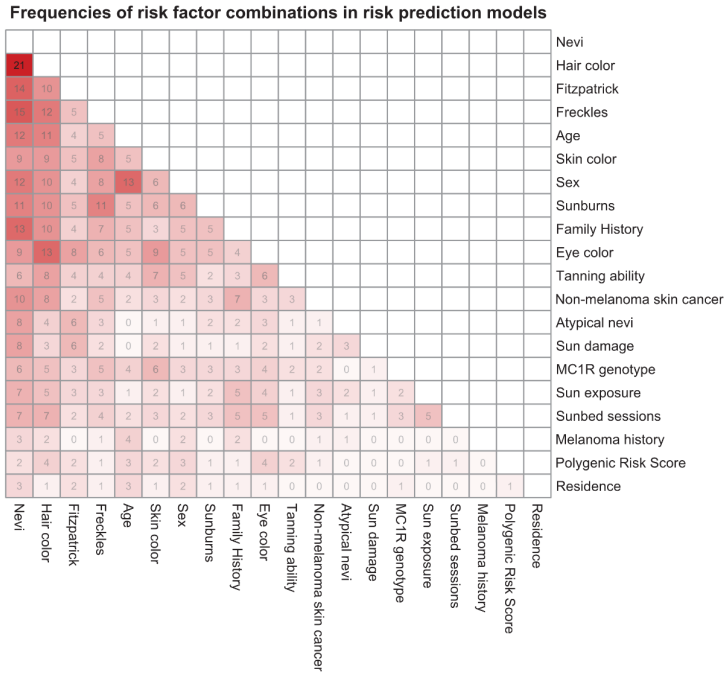

Figure 4. Heatmap indicating joint occurrences of risk factor pairs in risk prediction models for melanoma. Only risk factors occurring in more than two risk prediction models are included. Each number represents the absolute frequency of the corresponding risk factor combination. The darker the field the more frequent is the corresponding risk factor combination (n = 45 models). MC1R = melanocortin 1 receptor.

As studies used different definitions for at first glance identical predictors and employed varying methods of ascertaining them, Tables 2 and 3 show the heterogeneity in the definition and assessment of risk factors for the two examples, nevi and sunburns. Altogether, 69% of the models using nevi as predictor did not specify the size of nevi that were counted. No uniform procedure was also discernible for the body site on which the nevi were counted. Besides on the entire body (n = 16, 47%), nevi were also counted on only one or both arms and/or the back. Additionally, self-assessment of nevus counts by the study participant was approximately as frequent as the assessment by a professional, e.g., nurse or physician. Apart from four models, all others included nevi as a categorical variable. However, each study defined different categories (e.g., none/few/some/many or none/1-2/3-5/6-9/10+). Concerning the risk factor sunburns, six out of 13 studies did not specify how they define sunburns. The most frequent definition was "blistering," which was used by four studies. Furthermore, five studies asked for sunburns throughout the subject's lifetime, just as many addressed only sunburns in childhood and three studies did not specify this aspect. The reported measurement level of the variable was either dichotomous (n = 8, 61.5%) or categorical (n = 5, 38.5%).

Table 2. Heterogeneity of the risk factor nevi in risk prediction models for melanoma regarding four aspects: minimum size of nevi to be counted, body area of nevi count, type of nevi assessment, and measurement level (n = 35 models).

Risk Factors	*n*	%
Size of nevi		
≥2 mm	7	20.0
≥5 mm	2	5.7
>3 mm	1	2.9
≥2 mm and ≥5 mm, respectively	1	2.9
Not reported	24	68.6
Site of nevi count [(1)]		
Entire body	17	48.5
Both arms	6	17.1
Right arm	2	5.7
Forearm and back	2	5.7
Back	2	5.7
Left arm	1	2.9
Not reported	6	17.1
Assessment		
Physician/nurse/trained examiner	15	42.9
Self-reported	13	37.1
Not reported	7	20.0
Measurement level		
Categorical	31	88.6
Metric	2	5.7
Dichotomous	1	2.9
Not reported	1	2.9

[(1)] One model with nevi counted on two different sites.

Table 3. Heterogeneity of the risk factor sunburns in risk prediction models for melanoma regarding three aspects: definition of sunburn, time period of sunburn occurrence and measurement level (n = 13 models).

Category	*n*	%
Definition of sunburn		
Blistering	4	30.8
Pain and erythema or blisters for >24 h	1	7.7
Painful	1	7.7
Peeling of skin	1	7.7
No explanation given	6	46.2
Time period		
Childhood	5	38.5
Lifetime	5	38.5
Not reported	3	23.1
Measurement level		
Dichotomous	8	61.5
Categorial	5	38.5

3.4. Validation and Model Performance

In total, 18 out of all 40 studies used internal data to validate their models and only six used an external data set (Table 4), including three studies that did both internal and external validation. The remaining 19 studies used neither internal nor external validation. One of the studies with external validation used multiple external data sets, while the other five studies used just one external data set. For analyzing the temporal effect on the validation methods, we compared the two subgroups ("studies published up to 2011" and "studies published after 2011"). We found that the proportion of internal validations in the first subgroup is only one quarter compared to more than 60% in the subgroup of more recent studies. Furthermore, 70% of the older studies did not report any validation, while among the studies published after 2011 this proportion decreased to 25%.

Table 4. Absolute (n) and relative frequencies (%) of methods used when evaluating risk prediction models for melanoma regarding the methodological type of validation and the type of measures describing model performance (n = 40 studies) *.

	Studies Published up to 2011 (n = 20)		**Studies Published after 2011 (n = 20)**		**All Studies (n = 40)**	
Validation	***n***	**%**	***n***	**%**	***n***	**%**
Internal validation	**5**	**25.0**	**13**	**65.0**	**18**	**45.0**
Cross-validation	1	5.0	5	25.0	6	15.0
Split sample	3	15.0	3	15.0	6	15.0
Bootstrapping	1	5.0	5	25.0	6	15.0
External validation	**1**	**5.0**	**5**	**25.0**	**6**	**15.0**
Both internal and external validation	0	0.0	3	15.0	3	7.5
Neither internal nor external validation	14	70.0	5	25.0	19	47.5
Performance measures	***n***	**%**	***n***	**%**	***n***	**%**
Calibration (1)	**2**	**10.0**	**9**	**45.0**	**11**	**27.5**
Hosmer–Lemeshow test	2	10.0	7	35.0	9	22.5
Graph (plot/intercept/slope)	0	0.0	3	15.0	3	7.5
Calibration in the large	0	0.0	1	5.0	1	2.5

Table 4. *Cont.*

	Studies Published up to 2011 (n = 20)		Studies Published after 2011 (n = 20)		All Studies (n = 40)	
Discrimination [(2)]	**9**	**45.0**	**20**	**100.0**	**29**	**72.5**
AUC	8	40.0	18	90.0	26	65.0
C-index	0	0.0	3	15.0	3	7.5
Discrimination slope	0	0.0	1	5.0	1	2.5
ROC plot (without AUC calculation)	1	5.0	0	0.0	1	2.5
Overall model performance [(3)]	**0**	**0.0**	**1**	**5.0**	**1**	**2.5**
Brier score	0	0	1	5.0	1	2.5
Nagelkerk's R^2	0	0	1	5.0	1	2.5
Reclassification [(4)]	**0**	**0.0**	**4**	**20.0**	**4**	**10.0**
Net reclassification improvement	0	0.0	4.0	20.0	4	10.0
Integrated discrimination index	0	0.0	2.0	10.0	2	5.0
Clinically usefulness	**3**	**15.0**	**8**	**40.0**	**11**	**27.5**
Sensitivity/specificity	3	15.0	5	25.0	8	20.0
Decision curve	0	0.0	3	15.0	3	7.5
No performance measure at all	**11**	**55.0**	**0**	**0.0**	**11**	**27.5**

* For extended table with all references see Supplementary Table S2. [(1)] Two studies reported multiple calibration measures. [(2)] Two studies reported multiple discrimination measures. [(3)] One study reported both performance measures. [(4)] Two studies reported both reclassification measures.

Calibration was assessed in 11 studies and discrimination in 29 studies. The area under the curve (AUC) was the most common measure of discriminative performance (n = 26 studies, 65%). The proportion of studies published after 2011 reporting the AUC is 90%. Only one study reported measures for the overall model performance. Performance measures related to reclassification and clinical usefulness were less common (n = 4 and n = 11, respectively). In total, 73% of the studies reported a performance measure. Nine of the 20 studies published up to 2011 did not provide a measure of performance, while in the group of more recent studies, each study reported at least one performance measure.

4. Discussion

This systematic review shows that an abundance of melanoma risk prediction models were developed over the last decades, which has already been indicated by the systematic reviews of Vuong et al. [15] and Usher-Smith et al. [13] in 2014. We identified 40 studies reporting 46 risk prediction models that showed substantial heterogeneity in the choice of predictive factors and their definitions. In addition, only little consistency in model evaluation and poor validation were found among the studies.

The substantial heterogeneity of risk factors included in the prediction models is recognizable by the fact that 35 different risk factors were used in the risk prediction models. Whereas only some of them were frequently used, many (n = 15) were only used in one or two models. Six of the top 10 most common predictors were the phenotypic factors like nevi, hair color, Fitzpatrick skin type, freckles, skin color, and eye color, with nevi most frequently appearing in three quarters of the models. Thus, phenotypic predictors are currently dominating melanoma risk prediction, which could be traced back to their simple and fast assessment. Genetic factors, whose determination is more complex, are clearly less frequently used. Although polygenic risk scores may have the potential to meaningfully improve the predictive value of risk prediction models, their efficacy in clinical use is unproven [64].

As the heatmap in Figure 4 illustrates, risk factors were combined in various ways across all risk factor groups and only a few factor combinations were frequent. In fact, not even two of all melanoma risk prediction models in our analysis used exactly the same predictors. This heterogeneity of variables may be conditioned by the diversity of countries from where the studies originate. Nevertheless,

it complicates the validation by external data sets and direct comparisons between models on the same data set. Existing cohort studies may, for example, not be used for validation due to lack of information on one or more of the variables in the prediction model, and an existing data set used to develop one model does probably not contain all necessary predictors for other models. New data sets explicitly collected for preselected risk prediction models would be necessary.

By means of the two examples, nevi and sunburns, we demonstrated the considerable variation in the definition and assessment of risk factors, which only seem to be identical at first glance, as well as, in many cases, the lack of details in definitions and assessment methods given in the publications. About 70% of the models using nevi as risk factor did not specify the size of nevi counted. English and MacLennan published an IARC protocol for the identification and reporting of pigmented lesions like nevi and atypical nevi in 1990 [65]. Nevertheless, only the publication of Fortes et al. [23] cited the IARC protocol as reference for their nevi assessment, although 28 of the 35 risk prediction models with nevi as risk factor were developed after its publication. Regarding sunburns, a detailed definition enabling the distinction between a mild erythema and a painful sunburn was missing in almost half of the models. However, even if the definition is the same in several models, there are still many differences e.g., in the measurement levels or the categorization of quantitative variables. The more subjective a predictor is, the more variation is seen. This makes the comparability of the risk prediction models even more difficult. In order to externally validate models on independent data sets, uniform methods of data acquisition and a complete reporting of all information are necessary. One recent approach to the problem of non-uniform methods for data collection is given by the MelaNostrum consortium [66], which is a collaboration of researchers and clinicians from Mediterranean countries. They developed a consensus questionnaire of epidemiologic and clinical variables for melanoma risk assessment in order to standardize data collection across different studies, centers and languages [66].

Another significant cause for the heterogeneity of results and variables used in the reviewed studies could be traced back to the use of suboptimal primary data especially when attempting to capture individual UV-exposure over longer periods and behavioral aspects of UV protection. Self-reported sun exposure is commonly used, but it is uncertain how well this correlates with the actual sun exposure [67]. Biases from recalling past sun-related behavior and socially desirable answers are known problems, potentially limiting validity and reliability of exposure assessment [68–70]. Even when focusing on sunburn history, a commonly used marker of excessive UV exposure that has originally been believed to be better recalled, the ascertainment in self-administered questionnaires has been shown to lack reproducibility [71] and to underestimate the true extent of sunburns [72]. The validation of questionnaires would provide evidence for the magnitude of these problems, but this kind of validation is rare [67,68,73].

The analysis of the evaluation methods employed in the studies showed that the actual standard of validation and reporting of performance measures is quite poor. Nearly half of the studies (n = 18) validated their models internally using methods like bootstrapping and cross-validation, but only six out of 40 studies externally validated their models. The direct comparison of older and more recent studies reveals, however, an encouraging temporal effect. The proportion of studies without validation is clearly lower in the subgroup of studies published after 2011 compared to the subgroup of studies published up to 2011 (25% versus 70%). Nevertheless, there are still several exceptions among the newer studies [16,39,46,53,58], which did not report to have performed any validation of their model. Internal validation is useful for ensuring reproducibility of the model and stability of the predictor selection, as well as avoiding overfitting [47,74]. Applied to the data set used for developing the model, prediction models often perform substantially better than on external data sets, which can lead to overestimation of the predictive ability [75,76]. However, the models' performance on independent samples is a particularly valuable indicator for their discriminatory ability when applied in clinical practice [76]. Models developed for one population may not be valid in another population, as specific predictors do not necessarily apply in various regions. Therefore it is important to externally validate risk prediction models in order to verify generalizability and transportability

of the model to other cohorts [47]. To reduce overoptimistic expectations of model performance on independent data, Steyerberg and Harrell [77] proposed internal-external validation procedures during model development.

Considering the increasing number of different prediction models, studies about the external validation of multiple models may be the best way for a direct comparison of the models and to determine which model performs best on an independent data set [75]. Due to missing information about the model development and risk factor assessment in many cases, as discussed in previous sections, external validations particularly by independent researchers are rare [74]. One positive exception constitutes the study of Vuong et al. [49], who externally validated their risk prediction model using four independent data sets from other studies.

Further discrepancies among the studies were found concerning the performance measures. More than one-quarter of the publications (n = 11) did not report any performance measures, although a quantification of model performance is essential to evaluate if the model is suitable for application [48]. However, we observed an encouraging temporal effect concerning the evaluation of model performance too. For all categories of performance measures the proportion of studies reporting this kind of performance measure is higher in the subgroup of more recent studies than in the subgroup of older ones. Widely used statistical concepts for the evaluation of risk prediction models and the quality of prediction are discrimination and calibration. For the identification of individuals with a high risk of developing a certain disease, the discriminative ability is the most important property of the prediction models [78]. Almost 75% of the studies reported measures for the discriminative ability e.g., the AUC or c-index. Good calibration and discrimination are necessary but not sufficient for clinical usefulness. Hence, measures regarding the clinical usefulness can be regarded as essential to determine whether the model is beneficial for clinical practice and decision making [75,78]. Such performance measures related to clinical usefulness only emerged in the past few years, which may be the reason for their sparsity among the risk prediction models included in our analysis.

Even though this review relates to risk prediction models for melanoma, the problems identified are not limited to this area of medicine, as demonstrated by other publications of multivariable prediction or prognostic models addressing cardiovascular diseases, colorectal cancer, and diabetes type 2 [79–83]. Their results confirm the lack of validation and the need for more uniform assessment methods and independent validation. Due to the current focus on risk prediction models in general, it is an interdisciplinary problem that significantly more models have newly been developed than existing models were validated [74,75]. Furthermore, the fact that many risk factors are only used in one or two models indicates that the primary focus of most studies has been the identification of new predictors and the development of new risk prediction models, instead of the improvement and validation of existing models [79].

Complete reporting of how a prediction model has been developed and validated is necessary to objectively appraise the usefulness of the model [83]. Therefore it is a general requirement for all studies developing risk prediction models to report key details on model performance and their development process. The "Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis" (TRIPOD) statement [84], published in 2015, provides a checklist of 22 items essential for transparent reporting of a prediction model study. We found that only three of 10 studies published after the TRIPOD statement, reported their methods and results according to the statement and cited it as reference.

Considering the insufficient evaluation and the lack of quantitative comparisons, it is not surprising that in clinical practice no risk prediction model has been fully established in its entirety. Instead the individual risk is still mostly estimated based on single risk factors like count of nevi (>100 melanocytic nevi), congenital nevi (>40 cm in diameter) and Fitzpatrick skin type [85]. However, not all risk prediction models have been developed for clinical practice, e.g., the studies Barbini et al. [17], Whiteman and Green [35] and Fang et al. [59] are preliminary studies. Other recent examples are the

melanoma risk prediction models of Olsen et al. [52] and Vuong et al. [49], which are intended for community use in Australia as online risk calculators [86,87].

There are some limitations of our work that require consideration. First of all, the identification of studies reviewed here did not follow the PRISMA standard [88] and thus may not have the same degree of systematicity and objectivity. We performed, however, a comprehensive and careful, well-documented search for melanoma risk prediction models and identified several studies published after the systematic reviews of Vuong [15] and Usher-Smith [13] which were the starting points of our search. In fact, we found 16 publications of new melanoma risk prediction models that were published in the past six years. This relatively high number confirms the topicality of melanoma and the global interest in its prevention. A further limitation relates to the missing comparison of discriminative ability and overall performance of the different prediction models. We intentionally refrained from addressing this topic due to its methodologic complexity, which cannot be solved based on the information given in the study publications. Due to our abandonment of comparing the models with respect to their properties in practical applications, we also dropped the assessment of risk of bias for the studies having developed the prediction models. Shortcomings in study design, conduct and analysis can cast doubts on the study results and should be taken into consideration when evaluating prediction models comparatively in systematic reviews. Recently, the PROBAST tool has been developed through a consensus process to assess risk of bias and applicability of prediction model studies [89]. For our review we avoided using the PROBAST tool since its intention was beyond the scope of our project.

5. Conclusions

In summary, we found substantial heterogeneity in several important aspects of published risk prediction models for melanoma. Although a large number of models has been published, external validation is largely missing and direct comparisons between models are hardly possible. Consequently, there is no consensus in sight how to predict individual melanoma risk appropriately. Uniform standards for the assessment and documentation of predictors, as well as better adherence to reporting guidelines like TRIPOD are necessary and need to be obligatory in order to ultimately achieve a convincing solution of this problem.

Supplementary Materials: The following are available online at http://www.mdpi.com/1660-4601/17/21/7919/s1, File 1: Complete search strategy; Table S1: Classification of model variables into meaningful groups, Table S2: Absolute (n) and relative (%) frequencies of methods used when evaluating risk prediction models for melanoma regarding the methodological type of validation and the type of measures describing model performance (N = 40 studies). With references (ref.) Figure S1: Heatmap indicating relations joint occurrences of risk factor pairs among in risk prediction models for melanoma.

Author Contributions: Conceptualization, O.G.; methodology, I.K., A.B.P., W.U. and O.G.; software, I.K.; validation, O.G.; formal analysis, I.K.; investigation, I.K.; resources, A.B.P. and O.G.; data curation, A.B.P.; writing—original draft preparation, I.K.; writing—review and editing, A.B.P., W.U., M.V.H., M.B.V. and O.G.; visualization, I.K., M.V.H. and M.B.V.; supervision, A.B.P. and 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. Isabelle Kaiser performed the present work in partial fulfillment of the requirements for obtaining the degree 'Dr. rer.biol. hum'.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Basic characteristics of studies reporting risk prediction models for melanoma. Studies are ordered according to year of publication. (N = 40 studies).

Author (Year)	Study Design	Size of Study Sample	Year(s) of Data Collection	Country	Analytic Model	Risk Measure	Variables in Final Model [(1)]
English and Armstrong (1988) [20]	Case-control study	511 cases, 511 controls	1980–1981	Australia	Logistic regression	Risk score	Number of raised nevi on arms, age on arrival in Australia, mean time spent outdoors in summer aged 10–24, family history, personal history of non-melanoma skin cancer
Garbe et al. (1989) [25]	Case-control study	200 cases, 200 cases	1987	Germany	Logistic regression	Relative risks	Number of melanocytic common nevi, number of atypical nevi, actinic lentigines, occupational sun exposure, skin type
MacKie et al. (1989) [30]	Case-control study	280 cases, 280 controls	1987	Scotland	Logistic regression	Relative risk (risk groups)	Benign nevi >2 mm, freckling, atypical nevi >5 mm, episodes of severe sunburn
Augustsson et al. (1991) [40]	Case-control study	121 cases, 379 controls	1986–1988	Sweden	Logistic regression	Relative risks	Skin type, hair color, eye color, total body nevus ≥2 mm count, number of dysplastic nevi
Marret et al. (1992) [32]	Case-control study	583 cases, 608 controls	1984–1986	Canada	Logistic regression	Relative risk	Hair color, skin reaction to repeated sun exposure, freckle density, nevi density
Garbe et al. (1994) [24]	Case-control study	513 cases, 498 controls	1990–1991	Germany	Logistic regression	Relative risk estimates (risk groups)	Number of melanocytic common nevi, actinic lentigines, atypical nevi, skin type
Barbini et al. (1998) [17]	Case-control study	150 cases, 546 controls	1992–1995	Italy	Linear discriminant analysis	Risk score (negative score → low risk)	Colorimetric variables, Fitzpatrick
Landi et al. (2001) [29]	Case-control study	183 cases, 179 controls	1994–1999	Italy	Logistic regression	Odds ratios	Dysplastic nevi, skin color, tanning ability, eye color
Harbauer e al. (2003) [28]	Case-control study	202 cases, 202 controls	2001	Austria	Logistic regression	Odds ratios	Skin type, UV damage, number of nevi

Table A1. *Cont.*

Author (Year)	Study Design	Size of Study Sample	Year(s) of Data Collection	Country	Analytic Model	Risk Measure	Variables in Final Model [(1)]
Dwyer et al. (2004) [19]	Case-control study	244 cases, 483 controls	1998–1999	Australia	Logistic regression	Odds ratios	MC1R genotype, melanin density
Fargnoli et al. (2004) [21]	Case-control study	100 cases, 200 controls	2000–2001	Italy	Logistic regression	Relative risk estimates (high risk: ≥ 2 (3) risk factors in model 3 (1))	**Model 1**: Hair color, eye color, skin type Model 2: Hair color, eye color, skin type, occupational sun exposure, atypical nevi **Model 3**: Skin type, sun exposure, nevi, atypical nevi Model 4: Skin type, occupational sun exposure, nevi, atypical nevi
Cho et al. (2005) [18]	Cohort study	535 cases, total 178,155	1976, 1986, 1989	United States	Gail method	Risk score and 10-years-absolute risk	Sex, age, family history, sunburns, number of nevi on arms, hair color
Whiteman and Green (2005) [35]	Published case-control studies	NA	NA	Several countries	Not reported	10-years absolute risk	Age, place of residence, number of melanocytic nevi, skin color, MC1R genotype
Fears et al. (2006) [22]	Case-control study	718 cases, 945 controls	1991–1992	United States	Gail method	5-year-absolute risk (high risk: $p \geq 0.15\%$)	Sex, skin color, sunburns, number of moles >5 mm (only men), number of moles ≥2 mm, freckling, severe sun damage (only men), tanning ability (only women)
Goldberg et al. (2007) [26]	Cohort study	3329 cases, total 362,804	2001–2005	United States	Logistic regression	Risk score (high risk: score 4–5)	Sex, regular dermatologist, history of previous melanoma, mole changing, age
Fortes et al. (2010) [23]	Case-control study	304 cases, 305 controls	2001–2003	Italy	Risk score was calculated using effect estimates from meta-analysis	Individual risk score (high risk: risk score ≥ 3)	Freckles in childhood, skin color, number of common nevi, hair color, sunburns in childhood
Mar et al. (2011) [31]	Published meta-analysis and registry data	NA	NA	Australia	Gail method	5-year-absolute risk	Common nevi, atypical nevi, freckles, hair color, family history, non-melanoma skin cancer, personal melanoma history

Table A1. *Cont.*

Author (Year)	Study Design	Size of Study Sample	Year(s) of Data Collection	Country	Analytic Model	Risk Measure	Variables in Final Model [(1)]
Nielsen et al. (2011) [33]	Cohort study	215 cases, total 29,520	1990–1992 (Followup: –2007)	Sweden	Cox regression	Hazard ratios for each risk factor	Family history, number of nevi, hair color, sunbathing vacations, sunbed use
Quéreux et al. (2011) [37]	Case-control study	171 cases, 1390 controls	2007	France	Gail method, logistic regression and combinatorial analysis	Risk Score	Gail method: Sunburn in childhood, family history, number of common nevi on arms, density of freckles, skin type, recalled total sun exposure **Logistic regression** and combinatorial analysis: Sex, age, skin type, presence of freckles, number of nevi on arms, severe blistering sunburn in childhood, life in a country at low latitude, family history
Williams et al. (2011) [36]	Case-control study	386 cases, 727 controls	1997	United States	Logistic regression	Risk score (high risk: top 15%)	Sex, age, number of severe sunburns, hair color, freckles, number of raised moles, non-melanoma skin cancer history
Guther et al. (2012) [27]	Cohort study	250 cases, total 108,281	2005–2006	Germany	Logistic regression	Risk score (high risk: >0.0034)	Age, hair color, personal history of melanoma, suspicious melanocytic lesions
Smith et al. (2012) [39]	Case-control study	923 cases, 813 controls	Not reported	United States	Not reported	Not reported	Model A: Sex, age, hair color, eye color, mole count, freckling, family melanoma history **Model B**: Model A + outdoor UV, indoor UV, MC1R
Bakos et al. (2013) [16]	Case-control study	53 cases, 66 controls	2005–2008	Brazil	Risk score calculated using effect estimates from meta-analysis	Risk score (high risk: >3)	Presence of freckles in childhood, skin color, eye color, hair color, sunburn episodes throughout life

Table A1. *Cont.*

Author (Year)	Study Design	Size of Study Sample	Year(s) of Data Collection	Country	Analytic Model	Risk Measure	Variables in Final Model [1]
Cust et al. (2013) [45]	Case-control study	413 cases, 263 controls	2000–2002	Australia	Logistic regression	Odds ratios	Base model: Age, sex, city **Self-reported model**: MC1R genotype, nevi, pigmentation score [2], sun and sunbed exposure [3], family history, non-melanoma skin cancer **Physician-measured model**: Nevi, MC1R genotype, non-melanoma skin cancer, solar lentigines, family history, pigmentation score [4]
Fang et al. (2013) [59]	Multiple case-control studies	2298 cases, 6652 controls	NA	United States	Logistic regression	Odds ratios	Model 1: Single SNP Model 2: PRS [5] Model 3: Sex + age Model 4: Sex + age + pigmentation **Model 5**: Sex + age + pigmentation + PRS
Stefanaki et al. (2013) [34]	Case-control study	284 cases, 284 controls	NA	Greece	Logistic regression	Odds ratios	Model A: Eye color, hair color, skin color, skin type, tanning, sunburns Model B: All predictors in model A + 3 strongest SNPs **Model C**: All predictors in model A + all SNPs
Nikolic et al. (2014) [57]	Case-control study	341 cases, 356 controls	2001–2012	Serbia	Logistic regression + decision tree	Absolute risk	Level of education, intermitted exposure, use of sunbeds, HCT, solar damage of skin, Fitzpatrick, hair color, eye color, number of common nevi, number of dysplastic nevi, congenital nevi
Penn et al. (2014) [56]	Case-control study	875 cases, 765 controls	2004–2007	United States	Logistic regression	Odds ratios	Base model: Age, sex, hair color, eye color, skin color, freckles, mole phenotype **Full model**: Base model + sun burns, indoors tanning, MC1R genotype

Table A1. *Cont.*

Author (Year)	Study Design	Size of Study Sample	Year(s) of Data Collection	Country	Analytic Model	Risk Measure	Variables in Final Model [1]
Sneyd et al. (2014) [61]	Case-control study	368 cases, 270 controls	1992–1994	New Zealand	Logistic regression + Gail method	5-year-absolute risk	**Women:** Skin color, 1st degree relative with large or unusual moles, number of moles, personal history of nonmelanoma skin cancer **Men:** Number of moles, personal history, age at diagnosis, occupation, birthplace
Davies et al. (2015) [51]	Multiple case-control studies	NA	NA	Several countries	Logistic regression	Risk score (with risk categories)	Hair color, skin type, freckles, family history, nevi distribution, number of large nevi, sunburn
Kypreou et al. (2016) [54]	Case-control study	800 cases, 800 controls	2000–2014	Greece	Logistic regression	Odds ratios	Genetic risk score [6], age, sex, eye color, hair color, skin color, phototype, tanning ability
Vuong et al. (2016) [49]	Case-control study	629 cases, 535 controls	2000–2002	Australia	Gail method	20-year-absolute risk	Hair color, nevi density, family history, personal history of non-melanoma skin cancer, sunbed use
Cho et al. (2018) [58]	Cohort study	422/289 cases (lifetime/incident melanoma); total 19,102	NA–2015	United States	Logistic regression + Cox regression	Odds ratios and hazard ratios	Genetic risk score [7]
Cust et al. (2018) [62]	Case-control study	629 cases, 535 controls	2000–2002	Australia	Logistic regression	Odds ratios	Base model: Family history, hair color, nevi, personal history of non-melanoma skin cancer, sunburns in childhood, sunbed sessions, freckles, eye color, sun exposure **Full model:** Base model + PRS [8]
Gu et al. (2018) [60]	Case-control study	15,976 cases, 25,504 controls	NA	Several countries	Logistic regression	10- and 20-year-absolute risk	Model 1: Age, sex, country Model 2: (1) + eye color, hair color, skin type, common nevi Model 3: (1) + PRS [9] **Model 4:** (2) + PRS

Table A1. *Cont.*

Author (Year)	Study Design	Size of Study Sample	Year(s) of Data Collection	Country	Analytic Model	Risk Measure	Variables in Final Model [1]
Hübner et al. (2018) [53]	Cohort study based on data from SCREEN project	585 cases, total 354,635	2003–2004	Germany	Logistic regression	Odds ratios	Sex, age, personal melanoma history, family history, multiple common nevi, atypical nevi, congenital nevi
Olsen et al. (2018) [52]	Cohort study	655 cases, total 41,954	2011–2014	Australia	Cox regression	Hazard ratios	**Model 1 (invasive melanoma):** Age, sex, tanning ability, moles at age 21, hair color, number of previous skin lesions treated destructively, sunscreen use **Model 2 (all melanoma):** (1) + ethnicity, private health insurance, family history, past history of excisions for skin cancer, skin checks in past 3 years
Richter and Khoshgoftaar (2018) [55]	Cohort study based on EHR data	17,246 cases, total 9,531,408	2011–2017	United States	Logistic regression, decision tree + random forest	Risk score	Not reported
Tagliabue et al. (2018) [46]	Case-control study	3830 cases, 2619 controls	NA	Several countries	Logistic regression	Odds ratios	Base model: Age, sex, sunburns, number of common nevi, RH-phenotype **Base model + MC1R genotype**
Vuong et al. (2019) [50]	Case-control study	461 cases, 329 controls	2000–2002	Australia	Logistic regression	Relative risks	Number of nevi, solar lentigines, hair color, personal history of keratinocytic cancer

Abbreviations: UV = ultraviolet, MC1R = melanocortin 1 receptor, SNP = Single Nucleotide Polymorphism, PRS = Polygenic Risk Score, HCT = Hormonal Contraceptive Therapy, SCREEN = Skin Cancer Research to provide Evidence for Effectiveness of Screening in Northern Germany, EHR = Electronic Health Records, RH-Phenotype = Red Hair-Phenotype. [1] For studies with multiple models: Models included in analysis are highlighted in bold. [2] Score calculated from the variables: tanning ability, propensity to sunburn, skin color, eye color, hair color and freckles. [3] Term for the individual variables total childhood sun exposure, blistering sunburns and lifetime sunbed sessions. [4] Score was calculated from the following variables: hair color, eye color, skin reflectance, tanning ability, propensity to sunburn and freckles. [5] Comprised of 11 SNPs that demonstrated association with melanoma risk in previous studies. [6] Based on SNPs that showed genome-wide significant association with melanoma in previous studies. [7] Calculated using 21 genome-wide association study—significant SNPs. [8] Derived from 21 gene regions associated with melanoma. [9] Combines 204 common SNPs.

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.; Codon Publications: Brisbane, Australia, 2017.
2. Aitken, J.F.; Elwood, M.; Baade, P.D.; Youl, P.; English, D. Clinical whole-body skin examination reduces the incidence of thick melanomas. *Int. J. Cancer* **2010**, *126*, 450–458. [CrossRef] [PubMed]
3. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **2015**, *136*, E359–E386. [CrossRef] [PubMed]
4. Erdmann, F.; Lortet-Tieulent, J.; Schuz, J.; Zeeb, H.; Greinert, R.; Breitbart, E.W.; Bray, F. International trends in the incidence of malignant melanoma 1953–2008—Are recent generations at higher or lower risk? *Int. J. Cancer* **2013**, *132*, 385–400. [CrossRef] [PubMed]
5. International Agency for Research on Cancer. Available online: https://gco.iarc.fr/today/online-analysis-table?v=2018&mode=cancer&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=1&include_nmsc_other=1 (accessed on 23 July 2020).
6. Bertz, J. Epidemiologie des malignen Melanoms der Haut. *Bundesgesunheitsbl Gesundh. Gesundh.* **2001**, *44*, 484–490. [CrossRef]
7. Breitbart, E.W.; Waldmann, A.; Nolte, S.; Capellaro, M.; Greinert, R.; Volkmer, B.; Katalinic, A. Systematic skin cancer screening in Northern Germany. *J. Am. Acad Dermatol.* **2012**, *66*, 201–211. [CrossRef]
8. Boniol, M.; Autier, P.; Gandini, S. Melanoma mortality following skin cancer screening in Germany. *BMJ Open* **2015**, *5*, e008158. [CrossRef] [PubMed]
9. Halvorsen, J.A.; Loberg, M.; Gjersvik, P.; Roscher, I.; Veierod, M.B.; Robsahm, T.E.; Nilsen, L.T.N.; Kalager, M.; Bretthauer, M. Why a randomized melanoma screening trial is not a good idea. *Br. J. Dermatol.* **2018**, *179*, 532–533. [CrossRef]
10. U. S. Preventive Services Task Force; Bibbins-Domingo, K.; Grossman, D.C.; Curry, S.J.; Davidson, K.W.; Ebell, M.; Epling, J.W., Jr.; Garcia, F.A.; Gillman, M.W.; Kemper, A.R.; et al. Screening for Skin Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA* **2016**, *316*, 429–435.
11. Freedberg, K.A.; Geller, A.C.; Miller, D.R.; Lew, R.A.; Koh, H.K. Screening for malignant melanoma: A cost-effectiveness analysis. *J. Am. Acad Dermatol.* **1999**, *41*, 738–745. [CrossRef]
12. Watts, C.G.; Cust, A.E.; Menzies, S.W.; Mann, G.J.; Morton, R.L. Cost-effectiveness of skin surveillance through a specialized clinic for patients at high risk of melanoma. *J. Clin. Oncol.* **2017**, *35*, 63–71. [CrossRef]
13. Usher-Smith, J.A.; Emery, J.; Kassianos, A.P.; Walter, F.M. Risk prediction models for melanoma: A systematic review. *Cancer Epidemiol. Biomark. Prev.* **2014**, *23*, 1450–1463. [CrossRef] [PubMed]
14. Glanz, K.; Volpicelli, K.; Jepson, C.; Ming, M.E.; Schuchter, L.M.; Armstrong, K. Effects of Tailored Risk Communications for Skin Cancer Prevention and Detection: The PennSCAPE Randomized Trial. *Cancer Epidemiol. Biomark. Prev.* **2015**, *24*, 415–421. [CrossRef] [PubMed]
15. 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. [CrossRef]
16. 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. [CrossRef] [PubMed]
17. 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. [CrossRef]
18. 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. [CrossRef]
19. 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. [CrossRef]
20. English, D.R.; Armstrong, B.K. Identifying people at high-risk of cutaneous malignant-melanoma: Results from a case control study in Western Australia. *Brit. Med. J.* **1988**, *296*, 1285–1288. [CrossRef]

21. 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. [CrossRef]
22. Fears, T.R.; Guerry, D.; 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. [CrossRef]
23. 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. [CrossRef] [PubMed]
24. 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. Investig. Dermatol.* **1994**, *102*, 695–699. [CrossRef] [PubMed]
25. Garbe, C.; Kruger, S.; Stadler, R.; Guggenmoosholzmann, I.; Orfanos, C.E. Markers and relative risk in a German population for developing malignant-melanoma. *Int. J. Dermatol.* **1989**, *28*, 517–523. [CrossRef] [PubMed]
26. 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. [CrossRef]
27. 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.* **2012**, *26*, 86–94. [CrossRef]
28. 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. [CrossRef]
29. 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. [CrossRef]
30. MacKie, R.M.; Freudenberger, T.; Aitchison, T.C. Personal risk-factor chart for cutaneous melanoma. *Lancet* **1989**, *2*, 487–490. [CrossRef]
31. Mar, V.; Wolfe, R.; Kelly, J.W. Predicting melanoma risk for the Australian population. *Australas. J. Dermatol.* **2011**, *52*, 109–116. [CrossRef]
32. 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. *CMAJ* **1992**, *147*, 445–452.
33. 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. [CrossRef]
34. 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. [CrossRef] [PubMed]
35. Whiteman, D.C.; Green, A.C. A risk prediction tool for melanoma? *Cancer Epidemiol. Biomark. Prev.* **2005**, *14*, 761–763. [CrossRef] [PubMed]
36. 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*, 129.
37. 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. [CrossRef] [PubMed]
38. Weiss, J.; Garbe, C.; Bertz, J.; Blitz, H.; Burg, G.; Hennes, B.; Jung, E.G. Risk factors for the development of malignant melanoma in West Germany. Results of a multicenter-case control study. *Hautarzt* **1990**, *41*, 309–313.
39. Smith, L.A.; Qian, M.; Ng, E.; Shao, Y.; 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. [CrossRef]
40. Augustsson, A. Melanocytic naevi, melanoma and sun exposure. *Acta Derm. Venereol. Suppl.* **1991**, *166*, 1–34.
41. Zaridze, D.G.; Mukeriia, A.F.; Basieva, T.; Shlenskaia, I.N.; Bukin Iu, V. The role of endogenous and exogenous factors in the etiology of skin melanoma. *Vopr. Onkol.* **1992**, *38*, 141–147.

42. Wohlin, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE '14); Association for Computing Machinery: New York, NY, USA, 2014; pp. 1–10.
43. Felizardo, K.R.; Mendes, E.; Kalinowski, M.; Souza, É.F.; Vijaykumar, N.L. Using forward snowballing to update systematic reviews in software engineering. In Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement; Association for Computing Machinery: Ciudad Real, Spain, 2016.
44. Mendes, E.; Felizardo, K.; Wohlin, C.; Kalinowski, M. Search strategy to update systematic literature reviews in software engineering. In *2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*; Institute of Electrical and Electronics Engineers Inc.: Kallithea-Chalkidiki, Greece, 2019; pp. 355–362.
45. 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. [CrossRef]
46. 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. [CrossRef]
47. Steyerberg, E.W.; Vergouwe, Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur. Heart J.* **2014**, *35*, 1925–1931. [CrossRef] [PubMed]
48. Steyerberg, E.W.; Vickers, A.J.; Cook, N.R.; Gerds, T.; Gonen, M.; Obuchowski, N.; Pencina, M.J.; Kattan, M.W. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* **2010**, *21*, 128–138. [CrossRef]
49. 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. [CrossRef]
50. 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.* **2019**, *182*, 1262–1268. [CrossRef] [PubMed]
51. 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 Epidem. Biomar. Prev.* **2015**, *24*, 817–824. [CrossRef] [PubMed]
52. 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. Risk stratification for melanoma: Models derived and validated in a purpose-designed prospective cohort. *J. Natl. Cancer Inst.* **2018**, *110*, 1075–1083. [CrossRef]
53. Hübner, 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. [CrossRef]
54. 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. Investig. Dermatol.* **2016**, *136*, 690–695. [CrossRef]
55. Richter, A.; Khoshgoftaar, T. Melanoma risk prediction with structured electronic health records. In *ACM-BCB'18: 9th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*; Association for Computing Machinery: New York, NY, USA, 2018.
56. 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.
57. Nikolic, J.; Loncar-Turukalo, T.; Sladojevic, S.; Marinkovic, M.; Janjic, Z. Melanoma risk prediction models. *Vojnosanit. Pregl.* **2014**, *71*, 757–766. [CrossRef] [PubMed]
58. Cho, H.G.; Ransohoff, K.J.; Yang, L.Y.; Hedlin, H.; Assimes, T.; Han, J.L.; 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. [CrossRef] [PubMed]

59. Fang, S.Y.; Han, J.L.; Zhang, M.F.; Wang, L.E.; Wei, Q.Y.; Amos, C.I.; Lee, J.E. Joint Effect of Multiple Common SNPs Predicts Melanoma Susceptibility. *PLoS ONE* **2013**, *8*, e85642. [CrossRef] [PubMed]
60. Gu, F.Y.; 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. [CrossRef] [PubMed]
61. 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. [CrossRef] [PubMed]
62. Cust, A.E.; Drummond, M.; Kanetsky, P.A.; Goldstein, A.M.; Barrett, J.H.; MacGregor, S.; Law, M.H.; Iles, M.M.; Bui, M.; Hopper, J.L.; et al. Assessing the incremental contribution of common genomic variants to melanoma risk prediction in two population-based studies. *J. Investig. Dermatol.* **2018**, *138*, 2617–2624. [CrossRef] [PubMed]
63. World Cancer Research Fund. Available online: https://www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics (accessed on 6 March 2020).
64. Roberts, M.R.; Asgari, M.M.; Toland, A.E. Genome-wide association studies and polygenic risk scores for skin cancer: Clinically useful yet? *Br. J. Dermatol.* **2019**, *181*, 1146–1155. [CrossRef]
65. English, D.R.; MacLennan, R.; Rivers, J.; Kelly, J.; Armstrong, B.K. Epidemiological studies of melanocytic naevi protocol for identifying and recording naevi. In *IARC Internal Report No 90/002*; International Agency for Research on Cancer: Lyon, France, 1990.
66. Stratigos, A.J.; Fargnoli, M.C.; De Nicolo, A.; Peris, K.; Puig, S.; Soura, E.; Menin, C.; Calista, D.; Ghiorzo, P.; Mandala, M.; et al. MelaNostrum: A consensus questionnaire of standardized epidemiologic and clinical variables for melanoma risk assessment by the melanostrum consortium. *J. Eur. Acad Dermatol. Venereol.* **2018**, *32*, 2134–2141. [CrossRef]
67. Cargill, J.; Lucas, R.M.; Gies, P.; King, K.; Swaminathan, A.; Allen, M.W.; Banks, E. Validation of brief questionnaire measures of sun exposure and skin pigmentation against detailed and objective measures including vitamin D status. *Photochem. Photobiol.* **2013**, *89*, 219–226. [CrossRef]
68. Koster, B.; Sondergaard, J.; Nielsen, J.B.; Allen, M.; Olsen, A.; Bentzen, J. The validated sun exposure questionnaire: Association of objective and subjective measures of sun exposure in a Danish population-based sample. *Br. J. Dermatol.* **2017**, *176*, 446–456. [CrossRef]
69. Parr, C.L.; Hjartåker, A.; Laake, P.; Lund, E.; Veierød, 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.
70. Gefeller, O. Invited Commentary: Recall bias in melanoma—Much ado about almost nothing? *Am. J. Epidemiol.* **2009**, *169*, 267–270. [PubMed]
71. Veierød, M.B.; Parr, C.L.; Lund, E.; Hjartåker, A. Reproducibility of self-reported melanoma risk factors in a large cohort study of Norwegian women. *Melanoma Res.* **2008**, *18*, 1–9. [PubMed]
72. Koster, B.; Sondergaard, J.; Nielsen, J.B.; Olsen, A.; Bentzen, J. Reliability and consistency of a validated sun exposure questionnaire in a population-based Danish sample. *Prev. Med. Rep.* **2018**, *10*, 43–48.
73. 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. [PubMed]
74. Vickers, A.J. Prediction models in cancer care. *CA Cancer J. Clin* **2011**, *61*, 315–326. [PubMed]
75. Collins, G.S.; Moons, K.G. Comparing risk prediction models. *BMJ* **2012**, *344*, e3186.
76. 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.
77. Steyerberg, E.W.; Harrell, F.E., Jr. Prediction models need appropriate internal, internal-external, and external validation. *J. Clin. Epidemiol.* **2016**, *69*, 245–247.
78. Steyerberg, E.W. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*; Springer: New York, NY, USA, 2009.
79. Damen, J.A.; Hooft, L.; Schuit, E.; Debray, T.P.; Collins, G.S.; Tzoulaki, I.; Lassale, C.M.; Siontis, G.C.; Chiocchia, V.; Roberts, C.; et al. Prediction models for cardiovascular disease risk in the general population: Systematic review. *BMJ* **2016**, *353*, i2416.
80. Mallett, S.; Royston, P.; Waters, R.; Dutton, S.; Altman, D.G. Reporting performance of prognostic models in cancer: A review. *BMC Med.* **2010**, *8*, 21. [CrossRef] [PubMed]

81. Usher-Smith, J.A.; Walter, F.M.; Emery, J.D.; Win, A.K.; Griffin, S.J. Risk prediction models for colorectal cancer: A systematic review. *Cancer Prev. Res.* **2016**, *9*, 13–26. [CrossRef] [PubMed]
82. Mahar, A.L.; Compton, C.; Halabi, S.; Hess, K.R.; Gershenwald, J.E.; Scolyer, R.A.; Groome, P.A. Critical assessment of clinical prognostic tools in melanoma. *Ann. Surg. Oncol.* **2016**, *23*, 2753–2761. [CrossRef] [PubMed]
83. Collins, G.S.; Mallett, S.; Omar, O.; Yu, L.M. Developing risk prediction models for type 2 diabetes: A systematic review of methodology and reporting. *BMC Med.* **2011**, *9*, 103. [CrossRef]
84. 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. [CrossRef] [PubMed]
85. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF). S3-Leitlinie Prävention von Hautkrebs, Langversion 1. 1. 2014, AWMF Registernummer: 032/052OL. Available online: http://leitlinienprogramm-onkologie.de/Leitlinien.7.0.html (accessed on 9 September 2020).
86. QIMR Berghofer Medical Research Institute. Available online: https://publications.qimrberghofer.edu.au/Custom/QSkinMelanomaRisk (accessed on 12 September 2020).
87. Melanoma Institute Australia. Available online: https://www.melanomarisk.org.au/ (accessed on 12 September 2020).
88. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Group, P. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* **2009**, *339*, b2535. [CrossRef]
89. 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. [CrossRef]