



# Predicting keratinocyte carcinoma in patients with actinic keratosis: development and internal validation of a multivariable risk-prediction model

S. Tokez ,<sup>1</sup> M. Alblas ,<sup>2</sup> T. Nijsten ,<sup>1</sup> L.M. Pardo and M. Wakkee <sup>1</sup>

<sup>1</sup>Department of Dermatology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands

<sup>2</sup>Department of Public Health, Erasmus MC University Medical Centre, Rotterdam, the Netherlands

## Summary

### Correspondence

M. Wakkee

E-mail: m.wakkee@erasmusmc.nl

### Accepted for publication

14 December 2019

### Funding sources

No external funding.

### Conflicts of interest

None declared.

DOI 10.1111/bjd.18810

**Background** Patients with actinic keratosis (AK) are at increased risk for developing keratinocyte carcinoma (KC) but predictive factors and their risk rates are unknown.

**Objectives** To develop and internally validate a prediction model to calculate the absolute risk of a first KC in patients with AK.

**Methods** The risk-prediction model was based on the prospective population-based Rotterdam Study cohort. We hereto analysed the data of participants with at least one AK lesion at cohort baseline using a multivariable Cox proportional hazards model and included 13 a priori defined candidate predictor variables considering phenotypic, genetic and lifestyle risk factors. KCs were identified by linkage of the data with the Dutch Pathology Registry.

**Results** Of the 1169 AK participants at baseline, 176 (15.1%) developed a KC after a median follow-up of 1.8 years. The final model with significant predictors was obtained after backward stepwise selection and comprised the presence of four to nine AKs [hazard ratio (HR) 1.68, 95% confidence interval (CI) 1.17–2.42], 10 or more AKs (HR 2.44, 95% CI 1.65–3.61), AK localization on the upper extremities (HR 0.75, 95% CI 0.52–1.08) or elsewhere except the head (HR 1.40, 95% CI 0.98–2.01) and coffee consumption (HR 0.92, 95% CI 0.84–1.01). Evaluation of the discriminative ability of the model showed a bootstrap validated concordance index (c-index) of 0.60.

**Conclusions** We showed that the risk of KC in patients with AK can be calculated with the use of four easily assessable predictor variables. Given the c-index, extension of the model with additional, currently unknown predictor variables is desirable.

### What's already known about this topic?

- Patients with actinic keratosis (AK) are at increased risk of developing keratinocyte carcinoma (KC), including both squamous cell and basal cell carcinoma.
- However, risk rates and predictive factors are unknown and to date no risk-prediction model has been developed for patients with AK.

### What does this study add?

- We present a multivariable risk-prediction model with an additional tool to calculate the absolute risk of KC development in patients with AK.
- The number of AKs (4–9 or  $\geq 10$ ), location of AKs (upper extremity or elsewhere except head) and coffee consumption are significant predictors with a moderate discriminative ability.

- This prediction model may help in the management of patients with AK but extension with additional factors is desirable before clinical implementation.

Actinic keratoses (AKs) are premalignant lesions and can be considered a clinical biomarker for cutaneous photodamage.<sup>1</sup> Population-based studies report a high prevalence of AKs, especially in elderly people of European ancestry.<sup>2,3</sup> In the Netherlands, 23.5% of the population aged 50 years or older has one or multiple AKs.<sup>4</sup> Individual AKs may progress into cutaneous squamous cell carcinoma (cSCC). Additionally, as a marker of ultraviolet radiation (UVR)-induced DNA damage, the presence of AK is a risk factor for keratinocyte carcinoma (KC) in general, including basal cell carcinoma (BCC).<sup>5–7</sup> It is unclear which patients with AK will develop KCs and how high this risk rate is, although several AK characteristics such as the presence of multiple AKs and their anatomical site, as well as general phenotypic factors (e.g. light pigment status) and exposure-related items (e.g. high UVR exposure) have been described to increase progression risk.<sup>8–10</sup> Correctly identifying high-risk patients is important to detect KCs at an early stage and to ensure timely intervention. Moreover, stratified AK management may reduce patients' anxiety, provide better management for high-risk individuals, and optimize the use of healthcare resources.<sup>11</sup>

Until now, several KC prediction models have been developed regarding the occurrence of either a first or subsequent KC in the general population.<sup>12–15</sup> However, none of these assessed what factors predict a KC in an AK population, which is a very relevant question for many healthcare providers. We therefore aimed to develop a model to predict the absolute risk of a first KC in patients with AK, taking into account phenotypic, genetic and lifestyle risk factors, by analysing over 1000 participants with AK from the prospective population-based Rotterdam Study cohort (RS).

## Patients and methods

### Study population

The RS is a prospective population-based cohort study comprising 14 926 participants aged 45 years and older from the general population of Ommoord in Rotterdam, the Netherlands. From July 1989 to present, the participants have undergone regular examinations in a research facility and interviews are conducted at home about every 3–4 years. Between 2010 and 2016, complete skin examinations were performed during the RS routine, focusing on common skin diseases including AK as well as potential risk factors. We included participants with at least one AK lesion during one of these examinations in our model. The date of first AK diagnosis in the RS cohort served as the starting point of follow-up.

The RS has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02-1015) and

by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians. Details of the study design and objectives have been described before.<sup>16</sup>

### Case definition

The study outcome was defined as a first KC, either BCC or cSCC, after AK diagnosis. To identify all cases of KC, the RS participants were linked to the Dutch nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA) using encrypted patient data [combination of the patient's sex, birth date and first four to eight letters of the (maiden) family name]. Participants with a BCC or cSCC diagnosis prior to their AK diagnosis were excluded, as our study was focused on KC-naïve patients with AK. Follow-up of all participants ended at the time of KC diagnosis or, when this outcome measure was not met, at the date of censoring. Censoring events were death as assessed from the municipal register or end of available PALGA follow-up on 31 July 2018, whichever occurred first.

### Candidate predictor variables

The candidate predictor variables were selected a priori based on literature review and clinical expertise and were categorized as follows: AK-specific variables, phenotypic factors, lifestyle factors, UVR exposure variables and a genetic susceptibility variable.

As AK-specific variables, we included the number of AKs at diagnosis<sup>8,9,14,17–21</sup> (prespecified into 1–3, 4–9,  $\geq 10$  during skin examinations) and categorized the location of AKs into three main groups: head, upper extremities and elsewhere. In the case of AKs on multiple locations per participant, more than one location variable could be selected.

We included four phenotypic factors, namely age at AK diagnosis in the RS (years),<sup>12–14,17,22,23</sup> sex,<sup>12–14,17,18,22</sup> tendency to develop sunburn<sup>8,12–14,17,20,24,25</sup> and pigment status.<sup>8,18,24,26</sup> The latter constituted a combination of hair and eye colour when young, as reported previously.<sup>13</sup>

As lifestyle factors, smoking (never vs. current or ever)<sup>8,14,27,28</sup> and coffee consumption (cups per day)<sup>12,13,29,30</sup> were included. Regarding UVR exposure, we selected variables reflecting intermittent or chronic exposure to UVR. Intermittent UVR exposure<sup>8,20,23,31,32</sup> was defined as a combination of likeliness to be outdoors when the sun is shining/having mainly outside hobbies, going on holidays to a sunny country at least 4 weeks per year and sunbed usage of at least 10 times in the past 5 years. Chronic UVR

exposure<sup>8,17,18,20,26,27,33</sup> was assessed as the history of occupational outdoor work for at least 4 h per day during at least 25 years.

We calculated a genetic risk score (GRS) per patient with AK by retrieving seven single-nucleotide polymorphisms (SNPs) that were significantly associated with both BCC and cSCC occurrence from the most recent genome-wide association studies (Tables S1 and S2; see Supporting Information).<sup>34,35</sup> A detailed description of the GRS computation method is presented in Appendix S1 (see Supporting Information). In brief, a weighted GRS was calculated using the regression coefficients of published associations between the selected SNP and cSCC, which were similar for BCC.<sup>35</sup> The genetic scores were computed as follows:  $GRS = \sum \beta_i G_i$ ; where  $\beta_i$  is the log(odds ratio) of the SNP and  $G_i$  is the number of per SNP risk alleles (0, 1 or 2).

All predictor variables were measured at baseline, i.e. at the moment of AK diagnosis, DNA from whole blood was extracted at the start of each cohort (I–III) within RS. For lifestyle and UVR exposure variables, values from an earlier examination round were used if they were missing at baseline.

### Model development and performance

We used a Cox proportional hazards model to determine the probability of first KC development in patients with AK, taking censoring into account. Before starting the model development, collinearity among plausible categorical predictor variables was tested with Cramer's V statistic with no evidence found for multicollinearity. We imputed all missing predictor variables except for GRS 10 times using multivariate imputation by chained equations,<sup>36</sup> under the assumption that the data were missing at random. We included all candidate predictors, the outcome (KC or censored) and the follow-up time in years in the imputation model. Also, RS cohort number (I–III) and socioeconomic status of the participants were included as auxiliary variables.

Univariable analyses were performed for all candidate predictor variables and the occurrence of KC. For the continuous variables age and coffee consumption, we explored a possible nonlinear relationship using a natural cubic spline with two degrees of freedom. The use of a spline for these variables neither significantly improved the fit of our model (measured with the  $\chi^2$  value) nor provided graphical evidence for a nonlinear relationship. We therefore included these variables in their linear forms.

Regardless of their P-values in the univariable analyses, all candidate predictors were included in the multivariable model.<sup>37,38</sup> We reduced the multivariable model by backward stepwise selection using a two-sided statistical significance level of  $\alpha = 0.20$  as the cutoff point to reduce selection bias and optimism and to prevent the exclusion of important predictors.<sup>38</sup> The estimated regression coefficients and variances from the 10 imputed datasets were combined based on Rubin's rules.<sup>39</sup>

We assessed the predictive performance of our model in terms of discrimination using Harrell's concordance index (c-index). The c-index in survival context can be interpreted as the probability that the model assigns a higher predicted risk of KC development to a patient (from a randomly chosen pair of patients) that develops KC earlier in time compared with a patient developing KC later in time and varies from 0.5 (noninformative model) to 1.0 (perfect model).<sup>40</sup> As a means of internal validation, bootstrapping was used to correct the c-index for optimism.

To account for overfitting, we multiplied the regression coefficients from our final model with a shrinkage factor, which we estimated with bootstrapping (1000 replications). Shrinkage of regression coefficients towards average is meant to improve predictions in future patients by preventing extreme distributions of the predictions.<sup>38</sup>

A complete case analysis was performed as sensitivity analysis. Reporting of the model is done according to the TRIPOD statement.<sup>41</sup>

### Model presentation

To provide individualized predictions on the risk of first KC development in patients with AK, we made a risk-prediction tool based on the shrunk regression coefficients of our internally validated model using Microsoft Excel (2010).

Descriptive statistics were computed using IBM SPSS Statistics for Windows, version 24.0. (IBM Corp.; Armonk, NY, USA). Model development and internal validation were conducted using R statistical software version 3.5.0 (R Foundation for Statistical Computing; Vienna, Austria) with the mice, Hmisc and rms libraries.

## Results

### Study population

A selection of all participants with at least one AK lesion at baseline resulted in 1558 participants. After linkage with PALGA, 389 participants were excluded who had at least one KC prior to their AK diagnosis. The median follow-up of the remaining 1169 participants was 5.2 years [interquartile range (IQR) 3.5–6.9], during which 176 participants developed a KC at a median follow-up of 1.8 years (IQR 0.2–3.8). The majority of participants (58.9%) had one to three AK lesions at baseline, mainly located on the head (84.4%). The overall median age was 73.0 years (IQR 67.0–80.0) and 55% of all participants were men (Table 1).

### Predictors for a first keratinocyte carcinoma

In univariable analyses, the presence of four to nine AKs and 10 or more AKs, an AK localization outside the head or upper extremities and increasing age were significantly associated with a higher risk of KC development (Table 2). On the contrary, the risk of KC occurrence decreased per cup of coffee consumption [hazard ratio (HR) 0.92, 95% confidence interval (CI)

**Table 1** Descriptive characteristics of the 1169 participants with at least one actinic keratosis (AK) at baseline and cases of keratinocyte carcinoma (KC) (N = 176) separately

Candidate predictor variables	Category	Overall (N = 1169)	KC cases (N = 176)	Non-KC group (N = 993)
Number of participants		1169 (100%)	176 (15.1%)	993 (84.9%)
Follow-up time (years)	Median (IQR); (range)	5.2 (3.5–6.9); (0.0–7.9)	1.8 (0.2–3.8); (0.3–7.9)	5.7 (3.7–7.0); (0.0–7.3)
Age at AK diagnosis (years)	Median (IQR)	73.0 (67.0–80.0)	73.0 (67.0–79.0)	73.0 (67.0–80.0)
Sex	Male	643 (55.0%)	96 (54.5%)	547 (55.1%)
Number of AKs at diagnosis	1–3	689 (58.9%)	78 (44.3%)	611 (61.5%)
	4–9	290 (24.8%)	49 (27.8%)	241 (24.3%)
	≥ 10	190 (16.3%)	49 (27.8%)	141 (14.2%)
AK on the head <sup>a</sup>	No	182 (15.6%)	26 (14.8%)	156 (15.7%)
	Yes	987 (84.4%)	150 (85.2%)	837 (84.3%)
AK on upper extremities <sup>b</sup>	No	882 (75.4%)	132 (75.0%)	750 (75.5%)
	Yes	287 (24.6%)	44 (25.0%)	243 (24.5%)
AK on other locations <sup>c</sup>	No	973 (83.2%)	132 (75.0%)	841 (84.7%)
	Yes	196 (16.8%)	44 (25.0%)	152 (15.3%)
Pigment status <sup>d</sup>	Dark	222 (19.0%)	32 (18.2%)	190 (19.1%)
	Intermediate	618 (52.9%)	95 (54.0%)	523 (52.7%)
	Light	281 (24.0%)	43 (24.4%)	238 (24.0%)
	Missing	48 (4.1%)	6 (3.4%)	42 (4.2%)
Being easily sunburned	No	704 (60.2%)	100 (56.8%)	604 (60.8%)
	Yes	416 (35.6%)	69 (39.2%)	347 (34.9%)
	Missing	49 (4.2%)	7 (4.0%)	42 (4.2%)
Intermittent sun exposure <sup>e</sup>	No	114 (9.8%)	18 (10.2%)	96 (9.7%)
	Yes	732 (62.6%)	97 (55.1%)	635 (63.9%)
	Missing	323 (27.6%)	61 (34.7%)	262 (26.4%)
Outdoor work <sup>f</sup>	No	462 (39.5%)	74 (42.0%)	388 (39.1%)
	Yes	133 (11.4%)	20 (11.4%)	113 (11.4%)
	Missing	574 (49.1%)	82 (46.6%)	492 (49.5%)
Smoking	Never	357 (30.5%)	50 (28.4%)	307 (30.9%)
	Current or ever	798 (68.3%)	123 (69.9%)	675 (68.0%)
	Missing	14 (1.2%)	3 (1.7%)	11 (1.1%)
Coffee consumption (cups/day)	Median (IQR)	3.3 (1.4–3.3)	1.4 (1.4–3.3)	3.3 (1.4–3.3)
	Missing	131 (11.2%)	23 (13.1%)	108 (10.9%)
GRS	Median (IQR)	1.0 (1.0–1.1)	1.1 (1.0–1.1)	1.0 (1.0–1.1)
	Missing	159 (13.6%)	25 (14.2%)	134 (13.5%)

GRS, genetic risk score; IQR, interquartile range. <sup>a</sup>Presence of AK on the face, ears and/or scalp. <sup>b</sup>Presence of AK on the back of the hands and/or forearms. <sup>c</sup>Presence of AK on locations elsewhere (not specified). <sup>d</sup>A combination of hair and eye colour when young. <sup>e</sup>Combination variable of a confirmatory answer to one or more of the following questions:

- Are you likely to be outside when the sun is shining/do you mainly have outside hobbies?
- Do you go on holidays to a sunny country at least 4 weeks per year on average?
- Have you used a sunbed for at least 10 times during the past 5 years?

<sup>f</sup>To have been/worked outdoors for at least 4 h daily during at least 25 years.

0.84–1.01]. After backward stepwise selection, four predictor variables remained in the final model: number of AKs at diagnosis (either 4–9 or 10 or more), localization of AKs on the upper extremities, localization of AKs elsewhere except on the head, and coffee consumption. After adjustment for all other predictors in multivariable analysis, age was not significantly associated with KC anymore. Having 10 or more AKs was the strongest predictor with an almost 2.5 times higher hazard of KC development compared with the presence of one to three AKs (HR 2.44, 95% CI 1.65–3.61). Although evidence exists for a familial aggregation basis of skin cancer,<sup>34,35,42–44</sup> the GRS based on SNPs associated with KC did not increase the risk of KC development in our AK population.

A sensitivity analysis on 335 participants with no missing values yielded comparable HRs but without the AK location variables in the reduced multivariable model (Table S3; see Supporting Information). The overall apparent c-index of the final model was 0.61 (95% CI 0.56–0.66). After internal validation of the model with bootstrapping, the optimism corrected c-index reduced to 0.60 (95% CI 0.57–0.66).

### Model presentation

Figure 1 shows an image of the risk-prediction tool that can be used easily to predict an AK patient's risk of KC

**Table 2** Associations [hazard ratios (HRs) with confidence intervals (CIs)] between candidate predictor variables and development of a first KC (n = 176) using a Cox proportional hazards model

Candidate predictor variables	Coding	Univariable HR (95% CI)	Multivariable HR <sup>a</sup> (95% CI)
Age		1.01 (0.99–1.03)*	–
Sex	Female	1.03 (0.77–1.39)	–
Number of AKs at diagnosis	1–3	Reference	Reference
	4–9	1.59 (1.11–2.28)**	1.68 (1.17–2.42)**
	≥ 10	2.47 (1.73–3.53)***	2.44 (1.65–3.61)***
AK on the head <sup>b</sup>	Yes	1.09 (0.72–1.65)	–
AK on upper extremities <sup>c</sup>	Yes	0.99 (0.71–1.41)	0.75 (0.52–1.08)*
AK on other locations <sup>d</sup>	Yes	1.72 (1.23–2.43)***	1.40 (0.98–2.01)*
Pigment status <sup>e</sup>	Dark	Reference	–
	Intermediate	1.01 (0.68–1.51)	–
	Light	1.00 (0.63–1.57)	–
Being easily sunburned	Yes	1.11 (0.82–1.51)	–
Intermittent sun exposure <sup>f</sup>	Yes	0.84 (0.52–1.36)	–
Outdoor work <sup>g</sup>	Yes	0.93 (0.58–1.51)	–
Smoking	Ever	1.09 (0.78–1.51)	–
Coffee consumption (cups/day)		0.92 (0.84–1.01)*	0.92 (0.84–1.01)*
GRS		1.92 (0.58–6.31)	–

AK, actinic keratosis; GRS, genetic risk score. \*P-value < 0.20, \*\*P-value < 0.05, and \*\*\*P-value < 0.005. <sup>a</sup>Final model after backward step-wise selection. <sup>b</sup>Presence of AK on the face, ears and/or scalp. <sup>c</sup>Presence of AK on the back of the hands and/or forearms. <sup>d</sup>Presence of AK on locations elsewhere (not specified).

<sup>e</sup>A combination of hair and eye colour when young. <sup>f</sup>Combination variable of a confirmatory answer to one or more of the following questions: • Are you likely to be outside when the sun is shining/do you mainly have outside hobbies?

• Do you go on holidays to a sunny country at least 4 weeks per year on average?

• Have you used a sunbed for at least 10 times during the past 5 years?

<sup>g</sup>To have been/worked outdoors for at least 4 h daily during at least 25 years.

development, given the four prognostic factors from the final model. The regression coefficients of these predictors have been multiplied with an estimated shrinkage factor of 0.91. After filling in the individual values for each of these predictors, the tool calculates the percentage risk of a first KC in 1, 3 and 5 years. For example, a patient with 10 AKs spread over the upper extremity and other body sites except the head and who drinks three cups of coffee per day, has a 23% risk of KC development in 5 years. Formula File S1, an Excel spreadsheet containing this risk-prediction tool, is available for reference in the online Supporting Information.

## Discussion

Our population-based study with over 1000 participants provides the first risk-prediction model for an AK-specific patient group and encompasses readily available phenotypic, lifestyle, UVR and genetic KC susceptibility factors. The strongest predictor of a first KC was having 10 or more AKs at diagnosis, which increased the KC risk by almost 2.5-fold. This is in line with other cohort studies demonstrating a strong dose–response relationship between the number of AKs and the risk of a KC.<sup>9,19–21</sup> This finding could be explained through several theories. Firstly, cumulative UVR exposure underlies both AK and KC development. A study of the association between AKs and KCs showed that the aetiological factors for AK development were essentially equal to the aetiological factors for both

BCC and cSCC development.<sup>17</sup> Secondly, AKs can be seen as an early phase in the biological continuum that eventually culminates in cSCC, which means that some of the AKs in our cohort might have progressed directly to cSCC.<sup>21</sup> Thirdly, from the concept of field cancerization, the presence of multiple AKs forms the ultimate groundwork for the progression of epithelial carcinogenesis.<sup>45</sup>

Little is known about the risk of KC development based on AK affected body site. We found that AKs localized on the upper extremities significantly decreased and AKs localized outside the head and upper extremities significantly increased the risk of KC. This finding is consistent with a Dutch systematic review concluding that patients with AKs on the head or upper extremity regions are less likely to develop KCs compared with patients with AKs on the neck, trunk or lower extremities.<sup>46</sup> An explanation for our finding is not straightforward. It is remarkable that covered body sites showed higher risk rates than the more chronic sun-exposed head and upper extremity regions, which may hint to a different carcinogenesis pattern than chronic UVR exposure.

Coffee consumption is a much-discussed factor in the field of skin cancer carcinogenesis. In our analyses, we found that coffee consumption significantly reduced the risk of a first KC by 8% per cup of coffee. Findings from mainly laboratory and animal studies have indicated a possible protective effect of caffeine against KC development through induction of apoptosis in UVR-damaged keratinocytes as well as inhibition of



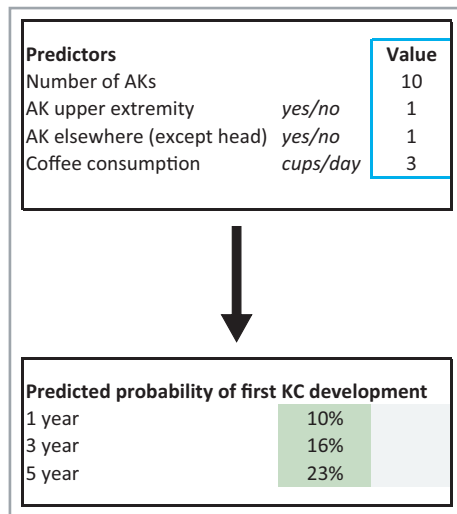


Fig 1. Risk-prediction tool for KC development in patients with AK, filled in for an example patient with 10 AKs, located on the upper extremity and elsewhere (not on the head), and who drinks three cups of coffee per day. The subsequent formula is used to predict the percentage risk of a first KC at 1 year after AK diagnosis:  $P = [1 - (\text{EXP}(-\text{EXP}(lp - lp_{\text{centered}}) * \text{baselinehaz}))] \times 100\%$  where  $lp = -0.278 * \text{AK location upper extremity} + 0.345 * \text{AK location elsewhere except head} - 0.060 * \text{cups of coffee per day} + \text{presence of multiple AKs}$  (0 if 1–3 AKs, 0.515 if 4–9 AKs, 0.888 if  $\geq 10$  AKs),  $lp_{\text{centered}} = 0.104$  and the baseline hazard is 0.057. Both  $lp$  and  $lp_{\text{centered}}$  have been multiplied by the shrinkage factor of 0.91. For the risks at 3 and 5 years, the baseline hazard should be replaced by 0.092 and 0.144, respectively.

UVR-induced carcinogenesis.<sup>47–49</sup> The chemo-protective effect of caffeine for KC (especially for BCC) in European-descent populations has recently been supported by two meta-analyses of observational studies as well.<sup>29,30</sup> Furthermore, the relation between coffee consumption and other malignancies has been investigated intensively: a significantly lower risk of cancers of, for example, the liver, endometrium, oral cavity and pharynx has been found.<sup>50–54</sup> Also, in these malignancies, the chemical and biological properties of coffee are mainly cited as the inducers of its positive effect. Additionally, we believe that coffee intake can be considered a proxy for good health and wellbeing as consumers of coffee often have a healthier lifestyle in general and therefore a lower risk of various malignancies, as argued by a recent review.<sup>55</sup> Focusing on the KC outcome, one could also hypothesize that people who drink more coffee are more often engaged in office jobs while people that rarely drink coffee are the ones involved in occupational outdoor work. This would result in a higher UVR exposure in the latter group and hence a higher KC risk, a potential effect which we were unable to adjust for (residual confounding).

Remarkably, none of the UVR-related predictor variables nor participants' pigment status was associated with KC. This is in line with other KC prediction models that used the same or comparable sun-exposure variables.<sup>7,13–15</sup> Because we selected our study population on the presence of AKs, which in a way can be considered primary KCs because of equal risk

profiles, index-event bias may underlie the results: UVR exposure is a pivotal risk factor for the occurrence of AKs, but in our model paradoxically not for a subsequent KC.<sup>56</sup> This is because conditioning on the presence of AKs generates dependence between all other known and unknown risk factors, eventually leading to underestimated or even reversed effects and biasing the risk rates towards the null. We indeed found HRs that were low (for being easily sunburned) or even seemed to be protective (history of outdoor work and intermittent sun exposure) in our univariable analyses, which are likely to be caused by index-event bias.

Regarding limitations, with the current internally validated discriminative value, our risk-stratification tool might not be clinically useful yet. Although we were able to include all variables of interest as derived from literature and clinical expertise, we found a c-index of 0.60. This poor-to-moderate c-index could be explained by the very homogeneous nature of our study population, which is an important distinction with prior models that were developed in a general population.<sup>14,15</sup> Patients with AK are a priori people with fair skin, at advanced age, and who have all had cumulative UVR exposure throughout the years. Finding additional KC predictors that specifically discriminate within the AK population is therefore a challenging task and the phenotypic, lifestyle and genetic risk factors at hand appeared to be insufficient. Another explanation for the moderate c-index might be that we have not separated BCC and cSCC as separate outcome measures due to insufficient power. Effect estimates per predictor could differ for BCC and cSCC, thereby influencing the discriminative ability of our model. However, a quick subgroup check on univariable analyses between the predictors and BCC/cSCC separately did not show any differences between both KC types (data not shown). Still, given the very limited existing knowledge in the AK prognostic field, we believe that the current model provides important insights and can be used to build on for more extensive models and the selection of tailored variables. Another limitation is that we assessed only the number of AKs at the moment of diagnosis during the RS, while this could have fluctuated during follow-up due to, for example, treatment or spontaneous regression of the lesion. Also, the number of AKs was already prespecified into the three categories during the skin examinations and we therefore could not include AK as a continuous variable in our model. However, as we assessed the overall risk of KC development considering all AKs in a patient instead of the lesion-specific progression risk, we do not expect that potential slight changes in the number of AKs would have affected the risk rates or the c-index of our model. Lastly, when interpreting our findings, one has to keep in mind that the study population comprised only people aged 45 years or older. Although AKs and KCs are mostly prevalent in the elderly population, this age criterion might limit the generalizability of our results. We have tried to find an independent cohort for external validation of our prediction model (QSkin Sun and Health Study from Australia).<sup>57</sup> Unfortunately, detailed information on AKs and other predictors from our model was not available.

In conclusion, the risk of first KC development in patients with AK can be predicted by a simple tool including the number and two location sites of AKs along with coffee consumption. This information can help physicians in identifying patients at high risk of KC and in planning further AK management. Extension with additional predictive factors and external validation thereafter are needed before use in clinical practice is recommended.

## Acknowledgments

The Rotterdam Study is funded by the Erasmus Medical Centre and Erasmus University Rotterdam; the Netherlands Organization for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European Commission (DG XII) and the Municipality of Rotterdam. We gratefully acknowledge the study participants and staff from the Rotterdam Study. We also thank Esther van den Broek from the Dutch Pathology Registry PALGA and Joris Verkouteren, Eline Noels and Lauren Onkenhout for their help with the linkage. We further thank Hester Lingsma, Daan Nieboer and Loes Hollestein for their statistical input.

## References

- Saladi RN, Persaud AN. The causes of skin cancer: a comprehensive review. *Drugs Today (Barc)* 2005; **41**:37–53.
- Naldi L, Chatenoud L, Piccitto R *et al.* Prevalence of actinic keratoses and associated factors in a representative sample of the Italian adult population: results from the Prevalence of Actinic Keratoses Italian Study, 2003–2004. *Arch Dermatol* 2006; **142**:722–6.
- Harvey I, Frankel S, Marks R *et al.* Non-melanoma skin cancer and solar keratoses II analytical results of the South Wales Skin Cancer Study. *Br J Cancer* 1996; **74**:1308–12.
- Flohil SC, van der Leest RJ, Dowlathshahi EA *et al.* Prevalence of actinic keratosis and its risk factors in the general population: the Rotterdam Study. *J Invest Dermatol* 2013; **133**:1971–8.
- Vatve M, Ortonne JP, Birch-Machin MA *et al.* Management of field change in actinic keratosis. *Br J Dermatol* 2007; **157** (Suppl. 2):21–4.
- Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. *Nat Rev Cancer* 2018; **18**:19–32.
- Footo JA, Harris RB, Giuliano AR *et al.* Predictors for cutaneous basal- and squamous-cell carcinoma among actinically damaged adults. *Int J Cancer* 2001; **95**:7–11.
- Belbasis L, Stefanaki I, Stratigos AJ *et al.* Non-genetic risk factors for cutaneous melanoma and keratinocyte skin cancers: an umbrella review of meta-analyses. *J Dermatol Sci* 2016; **84**:330–9.
- Khalesi M, Whiteman DC, Doi SA *et al.* Cutaneous markers of photo-damage and risk of basal cell carcinoma of the skin: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2013; **22**:1483–9.
- Richard MA, Amici JM, Basset-Seguín N *et al.* Management of actinic keratosis at specific body sites in patients at high risk of carcinoma lesions: expert consensus from the AKTeam of expert clinicians. *J Eur Acad Dermatol Venerol* 2018; **32**:339–46.
- Hollestein LM, de Vries E, Aarts MJ *et al.* Burden of disease caused by keratinocyte cancer has increased in the Netherlands since 1989. *J Am Acad Dermatol* 2014; **71**:896–903.
- Smedinga H, Verkouteren JAC, Steyerberg EW *et al.* Occurrence of metachronous basal cell carcinomas: a prognostic model. *Br J Dermatol* 2017; **177**:1113–21.
- Verkouteren JAC, Smedinga H, Steyerberg EW *et al.* Predicting the risk of a second basal cell carcinoma. *J Invest Dermatol* 2015; **135**:2649–56.
- Whiteman DC, Thompson BS, Thrift AP *et al.* A model to predict the risk of keratinocyte carcinomas. *J Invest Dermatol* 2016; **136**:1247–54.
- Wang W, Jorgenson E, Ioannidis NM *et al.* A prediction tool to facilitate risk-stratified screening for squamous cell skin cancer. *J Invest Dermatol* 2018; **138**:2589–94.
- Ikram MA, Brusselle GGO, Murad SD *et al.* The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol* 2017; **32**:807–50.
- Marks R, Rennie G, Selwood T. The relationship of basal cell carcinomas and squamous cell carcinomas to solar keratoses. *Arch Dermatol* 1988; **124**:1039–42.
- Oppel T, Korting HC. Actinic keratosis: the key event in the evolution from photoaged skin to squamous cell carcinoma. Therapy based on pathogenetic and clinical aspects. *Skin Pharmacol Physiol* 2004; **17**:67–76.
- Kricker A, Armstrong BK, English DR *et al.* Pigmentary and cutaneous risk factors for non-melanocytic skin cancer – a case-control study. *Int J Cancer* 1991; **48**:650–62.
- Armstrong BK, Kricker A. The epidemiology of UV induced skin cancer. *J Photochem Photobiol B* 2001; **63**:8–18.
- Salasche SJ. Epidemiology of actinic keratoses and squamous cell carcinoma. *J Am Acad Dermatol* 2000; **42**:4–7.
- Lovatt TJ, Lear JT, Bastrilles J *et al.* Associations between ultraviolet radiation, basal cell carcinoma site and histology, host characteristics, and rate of development of further tumors. *J Am Acad Dermatol* 2005; **52**:468–73.
- van der Geer S, Kleingeld PA, Snijders CC *et al.* Development of a non-melanoma skin cancer detection model. *Dermatology* 2015; **230**:161–9.
- Khalesi M, Whiteman DC, Tran B *et al.* A meta-analysis of pigmentary characteristics, sun sensitivity, freckling and melanocytic nevi and risk of basal cell carcinoma of the skin. *Cancer Epidemiol* 2013; **37**:534–43.
- Kennedy C, Bajdik CD, Willemze R *et al.* The influence of painful sunburns and lifetime sun exposure on the risk of actinic keratoses, seborrheic warts, melanocytic nevi, atypical nevi, and skin cancer. *J Invest Dermatol* 2003; **120**:1087–93.
- Green A, Battistutta D. Incidence and determinants of skin cancer in a high-risk Australian population. *Int J Cancer* 1990; **46**:356–61.
- Aubry F, MacGibbon B. Risk factors of squamous cell carcinoma of the skin. A case-control study in the Montreal region. *Cancer* 1985; **55**:907–11.
- Leonardi-Bee J, Ellison T, Bath-Hextall F. Smoking and the risk of nonmelanoma skin cancer: systematic review and meta-analysis. *Arch Dermatol* 2012; **148**:939–46.
- Caini S, Cattaruzza MS, Bendinelli B *et al.* Coffee, tea and caffeine intake and the risk of non-melanoma skin cancer: a review of the literature and meta-analysis. *Eur J Nutrition* 2017; **56**:1–12.
- Vaseghi G, Haghjoo-Javanmard S, Naderi J *et al.* Coffee consumption and risk of nonmelanoma skin cancer: a dose-response meta-analysis. *Eur J Cancer Prev* 2018; **27**:164–70.
- Kricker A, Armstrong BK, English DR *et al.* Does intermittent sun exposure cause basal cell carcinoma? A case-control study in Western Australia. *Int J Cancer* 1995; **60**:489–94.
- Wehner MR, Shive ML, Chren MM *et al.* Indoor tanning and non-melanoma skin cancer: systematic review and meta-analysis. *BMJ* 2012; **345**:e5909.

- 33 Schmitt J, Seidler A, Diepgen TL *et al.* Occupational ultraviolet light exposure increases the risk for the development of cutaneous squamous cell carcinoma: a systematic review and meta-analysis. *Br J Dermatol* 2011; **164**:291–307.
- 34 Chahal HS, Lin Y, Ransohoff KJ *et al.* Genome-wide association study identifies novel susceptibility loci for cutaneous squamous cell carcinoma. *Nat Commun* 2016; **7**:12048.
- 35 Sordillo JE, Kraft P, Wu AC *et al.* Quantifying the polygenic contribution to cutaneous squamous cell carcinoma risk. *J Invest Dermatol* 2018; **138**:1507–10.
- 36 Van Buuren S. *Flexible Imputation of Missing Data*. Boca Raton, FL, USA: Chapman and Hall/CRC, 2012.
- 37 Royston P, Moons KG, Altman DG *et al.* Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009; **338**:b604.
- 38 Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY, USA: Springer Science+Business Media, LLC, 2009.
- 39 Vergouwe Y, Royston P, Moons KG *et al.* Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol* 2010; **63**:205–14.
- 40 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; **15**:361–87.
- 41 Collins GS, Reitsma JB, Altman DG *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015; **350**:g7594.
- 42 Hussain SK, Sundquist J, Hemminki K. The effect of having an affected parent or sibling on invasive and in situ skin cancer risk in Sweden. *J Invest Dermatol* 2009; **129**:2142–7.
- 43 Hemminki K, Zhang H, Czene K. Familial invasive and in situ squamous cell carcinoma of the skin. *Br J Cancer* 2003; **88**:1375–80.
- 44 Chahal HS, Wu W, Ransohoff KJ *et al.* Genome-wide association study identifies 14 novel risk alleles associated with basal cell carcinoma. *Nat Commun* 2016; **7**:12510.
- 45 Torezan LA, Festa-Neto C. Cutaneous field cancerization: clinical, histopathological and therapeutic aspects. *An Bras Dermatol* 2013; **88**:775–86.
- 46 Smit P, Plomp E, Neumann HA *et al.* The influence of the location of the lesion on the absolute risk of the development of skin cancer in a patient with actinic keratosis. *J Eur Acad Dermatol Venerol* 2013; **27**:667–71.
- 47 Heffernan TP, Kawasumi M, Blasina A *et al.* ATR–Chk1 pathway inhibition promotes apoptosis after UV treatment in primary human keratinocytes: potential basis for the UV protective effects of caffeine. *J Invest Dermatol* 2009; **129**:1805–15.
- 48 Han W, Ming M, He YY. Caffeine promotes ultraviolet B-induced apoptosis in human keratinocytes without complete DNA repair. *J Biol Chem* 2011; **286**:22825–32.
- 49 Huang MT, Xie JG, Wang ZY *et al.* Effects of tea, decaffeinated tea, and caffeine on UVB light-induced complete carcinogenesis in SKH-1 mice: demonstration of caffeine as a biologically important constituent of tea. *Cancer Res* 1997; **57**:2623–9.
- 50 Alicandro G, Tavani A, La Vecchia C. Coffee and cancer risk: a summary overview. *Eur J Cancer Prev* 2017; **26**:424–32.
- 51 Zhang Y, Wang X, Cui D. Association between coffee consumption and the risk of oral cancer: a meta-analysis of observational studies. *Int J Clin Exp Med* 2015; **8**:11657–65.
- 52 Galeone C, Tavani A, Pelucchi C *et al.* Coffee and tea intake and risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol Biomarkers Prev* 2010; **19**:1723–36.
- 53 Turati F, Galeone C, La Vecchia C *et al.* Coffee and cancers of the upper digestive and respiratory tracts: meta-analyses of observational studies. *Ann Oncol* 2011; **22**:536–44.
- 54 Miranda J, Monteiro L, Albuquerque R *et al.* Coffee is protective against oral and pharyngeal cancer: a systematic review and meta-analysis. *Med Oral Patol Oral Cir Bucal* 2017; **22**:e554–e61.
- 55 Mirza SS, Tiemeier H, de Bruijn RF *et al.* Coffee consumption and incident dementia. *Eur J Epidemiol* 2014; **29**:735–41.
- 56 Dahabreh IJ, Kent DM. Index event bias as an explanation for the paradoxes of recurrence risk research. *JAMA* 2011; **305**:822–3.
- 57 Olsen CM, Green AC, Neale RE *et al.*; QSkin Study. Cohort profile: the QSkin Sun and Health Study. *Int J Epidemiol* 2012; **41**:929–929i.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Appendix S1** Genetic risk score (GRS) computation method.

**Table S1** Summary of single-nucleotide polymorphisms (SNPs) that have been associated with both basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (cSCC). Table presents the effect allele and the other allele for cSCC.

**Table S2** Characteristics of the SNPs used for the genetic risk score (GRS) within the RS.

**Table S3** Associations [hazard ratios (HRs) with confidence intervals (CIs)] between candidate predictor variables and development of a first KC (n = 48) in patients without any missing values (n = 335).

**Formula File S1** Interactive tool (in Excel spreadsheet) that can be filled in to calculate the absolute risk of KC in patients with AK.