A Model to Predict the Risk of Keratinocyte Carcinomas

David C. Whiteman[1], Bridie S. Thompson[1], Aaron P. Thrift[2], Maria-Celia Hughes[1], Chiho Muranushi[1,3], Rachel E. Neale[1], Adele C. Green[1,4] and Catherine M. Olsen[1], for the QSkin Study

Basal cell and squamous cell carcinomas of the skin are the commonest cancers in humans, yet no validated tools exist to estimate future risks of developing keratinocyte carcinomas. To develop a prediction tool, we used baseline data from a prospective cohort study (n = 38,726) in Queensland, Australia, and used data linkage to capture all surgically excised keratinocyte carcinomas arising within the cohort. Predictive factors were identified through stepwise logistic regression models. In secondary analyses, we derived separate models within strata of prior skin cancer history, age, and sex. The primary model included terms for 10 items. Factors with the strongest effects were >20 prior skin cancers excised (odds ratio 8.57, 95% confidence interval [95% CI] 6.73–10.91), >50 skin lesions destroyed (odds ratio 3.37, 95% CI 2.85–3.99), age ≥ 70 years (odds ratio 3.47, 95% CI 2.53–4.77), and fair skin color (odds ratio 1.75, 95% CI 1.42–2.15). Discrimination in the validation dataset was high (area under the receiver operator characteristic curve 0.80, 95% CI 0.79–0.81) and the model appeared well calibrated. Among those reporting no prior history of skin cancer, a similar model with 10 factors predicted keratinocyte carcinoma events with reasonable discrimination (area under the receiver operator characteristic curve 0.72, 95% CI 0.70–0.75). Algorithms using self-reported patient data have high accuracy for predicting risks of keratinocyte carcinomas.

Journal of Investigative Dermatology (2016) **136,** 1247–1254; doi:10.1016/j.jid.2016.02.008

INTRODUCTION

Keratinocyte carcinomas (KCs) (specifically, basal cell carcinomas (BCC) and squamous cell carcinomas (SCC) of the skin) are the most common cancers in humans. Although often regarded as unimportant cancers, KCs cause considerable morbidity. Each year, 1.9% of the US adult population is estimated to receive treatment for KC, rising to 6.9% of the population aged >65 years (U.S. Department of Health and Human Services, 2014). As such, KCs are the most costly cancers in the US Medicare population, contributing to the estimated total direct treatment costs of $4.3 billion each year (Housman et al., 2003; U.S. Department of Health and Human Services, 2014). Globally, the highest rates of BCC (>1000 × 10^{-5} person-years) and SCC (387 × 10^{-5} person-years) are observed in Australia, although KC incidence rates exceed 100 × 10^{-5} person-years in most fair-skinned populations around the world (Lomas et al., 2012).

The enormous burden of skin cancers underscores the need to find better ways to control these conditions, both through primary prevention and early detection (U.S. Department of Health and Human Services, 2014). Reducing hazardous exposure to solar ultraviolet radiation is accepted as the mainstay of primary prevention in the general population, but there is less certainty about how to deploy medical services for skin cancer control. At one end of the early detection spectrum is formal population-based screening. To date, only Germany has embarked upon a nationwide program of screening for skin cancer (Choudhury et al., 2012), although evidence is emerging that the benefits may not justify the costs (Katalinic et al., 2015; Stang and Jockel, 2015). In other jurisdictions, guidelines recommend that patients at high risk for skin cancer undergo periodic surveillance, with the implied but unstated advice that those at low risk receive usual care (Cancer Council Australia, 2007; U.S. Preventive Services Task Force, 2009). With the advent of promising chemoprevention strategies for skin cancer (Chen et al., 2015), reliable risk stratification tools are needed to identify patients most likely to benefit from clinical intervention. Presently, clinicians must rely on their own experience to estimate a patient's future risk of skin cancer, because unlike for melanoma (Olsen et al., 2015), no prediction tools for these cancers have been developed or validated. Thus, we aimed to develop and validate a risk prediction model for quantifying the probability of being treated for a KC, regardless of histological subtype. Our approach was to use self-reported information of the type that can be elicited remotely, so that patients might be assessed and triaged before consulting a physician.

RESULTS

The eligible cohort comprised 38,726 participants with no prior history of melanoma, of whom 56% were women and the mean age was 56.2 years (standard deviation 8.1). Most

[1]Population Health Department, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia; [2]Department of Medicine and Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas, USA; [3]School of Population Health, The University of Queensland, Herston, Queensland, Australia; and [4]Cancer Research UK Manchester Institute and Institute of Inflammation and Repair, University of Manchester, Manchester, UK

Correspondence: David C. Whiteman, Cancer Control Group, QIMR Berghofer Medical Research Institute, Locked Bag 2000, RBWH, Queensland 4029, Australia. E-mail: david.whiteman@qimrberghofer.edu.au

Abbreviations: AUROC, area under the receiver operator characteristic curve; BCC, basal cell carcinomas; KC, keratinocyte carcinoma; SCC, squamous cell carcinomas

Received 16 November 2015; revised 19 January 2016; accepted 1 February 2016; accepted manuscript published online 22 February 2016; corrected proof published online 22 March 2016

participants reported white European ancestry (n = 34,579, 93%) and most were born in Australia (30,054, 81%). Median follow-up was 36 months (min 20; max 44). Distributions of key factors in the derivation and validation datasets are presented in Table 1.

There were 4,237 (16%) participants in the derivation dataset with at least one surgical excision for a confirmed BCC or SCC during the follow-up period. In univariate analyses, 23 items were significantly associated with the occurrence of KC (Supplementary Table S1 online). After conducting backward stepwise regression within the imputed derivation datasets (Supplementary Table S2 online), and then testing whether adding further candidate terms significantly improved the fit of the model, we derived a final model with terms for 10 predictors (age, sex, smoking status, ethnicity, skin color, tanning ability, freckling tendency, number of sunburns <10 years, number of previous skin cancers excised, and number of previous skin lesions destroyed; Table 2). None of the pairwise interaction terms was statistically significant at the $P < 0.05$ level, nor did they substantially modify the Akaike Information Criterion or the area under the receiver operator characteristic curve (AUROC); hence they were not retained in the final models. Strongest effects were observed for >20 prior skin cancers excised, >50 skin lesions destroyed, age >70 years, and fair skin color. Discrimination in the validation dataset was high (AUROC 0.80, 95% confidence interval 0.79–0.81; Figure 1). Although the model appeared well calibrated overall, there was some evidence that true risks were underestimated at the low end of the scale, and that risks were marginally overestimated in the highest categories (Figure 2).

Cumulative incidence plots demonstrated that the risk of KC events was strongly predicted by the self-reported history of prior skin cancer excisions (Figure 3). Thus, we derived separate models according to the absence or presence of a past skin cancer history. Among those in the derivation dataset with no prior history of skin cancer excisions (n = 16,021), we derived a predictive model with good discrimination (AUROC 0.72, 95% confidence interval 0.70–0.75;

Table 1. Characteristics of study participants in derivation and validation datasets

	Derivation dataset			Validation dataset		
Characteristic	**Controls (N = 21,605)**	**KC cases (N = 4,237)**	**Total (N = 25,842)**	**Controls (N = 10,773)**	**KC cases (N = 2,111)**	**Total (N = 12,884)**
Age group						
40–49	6,309 (29.2)	616 (15.4)	6,925 (26.8)	3,136 (29.1)	273 (12.9)	3,409 (26.5)
50–59	8,371 (38.8)	1,471 (34.7)	9,842 (38.1)	4,183 (38.8)	722 (34.2)	4,905 (38.1)
60–69	6,784 (31.4)	2,063 (48.7)	8,847 (34.2)	3,377 (31.4)	1,069 (50.6)	4,446 (34.5)
70+	141 (0.7)	87 (2.1)	228 (0.9)	77 (0.7)	47 (2.2)	124 (1.0)
Sex						
Female	12,327 (57.1)	1,933 (45.6)	14,260 (55.2)	6,094 (56.6)	975 (46.2)	7,069 (54.9)
Male	9,278 (42.9)	2,304 (53.7)	11,582 (44.8)	4,679 (43.4)	1,136 (53.8)	5,815 (45.1)
Ethnicity						
Non-white	1,500 (6.9)	85 (2.0)	1,585 (6.9)	747 (6.9)	42 (2.0)	789 (6.1)
White	19,889 (92.1)	4,115 (97.1)	24,004 (92.9)	9,917 (92.1)	2,055 (97.4)	11,972 (92.9)
Missing	216 (1.0)	37 (0.9)	253 (1.0)	109 (1.0)	14 (0.7)	123 (1.0)
Born in Australia						
No	4,537 (21.0)	592 (14.0)	5,129 (21.0)	2,272 (21.1)	280 (13.3)	2,552 (19.8)
Yes	17,062 (78.0)	3,641 (85.9)	20,703 (78.0)	8,495 (78.9)	1,831 (86.7)	10,326 (80.2)
Missing	6 (<0.1)	4 (0.1)	10 (<0.1)	6 (0.1)	0 (0)	6 (0.1)
Private health insurance						
No	7,311 (33.8)	1,313 (31.0)	8,624 (33.4)	3,564 (33.1)	624 (29.6)	4,188 (32.5)
Yes	14,191 (65.7)	2,900 (68.4)	17,091 (66.1)	7,168 (66.5)	1,480 (70.1)	8,648 (67.1)
Missing	103 (0.5)	24 (0.6)	127 (0.5)	41 (0.4)	7 (0.3)	48 (0.4)
Education						
University degree	5,341 (24.7)	855 (20.2)	6,196 (25.0)	2,691 (25.0)	410 (19.4)	3,101 (24.1)
Certificate or diploma	4,070 (18.8)	691 (16.3)	4,761 (18.4)	2,145 (19.9)	328 (15.5)	2,473 (19.2)
Trade	1,954 (9.0)	414 (9.8)	2,368 (9.2)	963 (8.9)	213 (10.1)	1,176 (17.8)
Secondary school	3,974 (18.4)	736 (17.4)	4,710 (18.2)	1,906 (17.7)	386 (18.3)	2,292 (17.8)
Intermediate school	3,200 (14.8)	764 (18.0)	3,964 (15.3)	1,550 (14.4)	398 (18.9)	1,948 (15.1)
Primary school	1,575 (7.3)	417 (9.8)	1,992 (7.7)	785 (7.3)	194 (9.2)	979 (7.6)
Missing	1,491 (6.9)	360 (8.5)	1,851 (7.2)	733 (6.8)	182 (8.6)	915 (7.1)
Smoking status						
Never	11,763 (54.5)	2,260 (53.3)	14,023 (54.3)	5,944 (55.2)	1,155 (54.7)	7,099 (55.1)
Past	7,610 (35.2)	1,523 (36.0)	9,133 (35.3)	3,760 (34.9)	743 (35.2)	4,503 (35.0)
Current	2,120 (9.8)	424 (10.0)	2,544 (9.8)	1,002 (9.3)	203 (9.6)	1,205 (9.4)
Missing	112 (0.5)	30 (0.7)	142 (0.5)	67 (0.6)	67 (0.6)	77 (0.6)

Abbreviation: KC, keratinocyte carcinoma.