BMJ Open

Prediction of breast cancer risk among women of the Mariana Islands: the BRISK model

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-061205
Article Type:	Original research
Date Submitted by the Author:	18-Jan-2022
Complete List of Authors:	Shvetsov, Yurii; University of Hawai'i at Mānoa, Cancer Center Wilkens, Lynne ; University of Hawai'i at Mānoa, Cancer Center White, Kami; University of Hawai'i at Mānoa, Cancer Center Chong, Marie; University of Hawai'i at Mānoa, Cancer Center Buyum, Arielle; AB Consulting, LLC Badowski, Grazyna; University of Guam, College of Natural and Applied Sciences Leon Guerrero, Rachael; University of Guam, College of Natural and Applied Sciences Novotny, Rachel; University of Hawai'i at Manoa, College of Tropical Agriculture and Human Resources
Keywords:	Breast tumours < ONCOLOGY, Epidemiology < ONCOLOGY, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Prediction of breast cancer risk among women of the Mariana Islands: the BRISK model

Yurii B. Shvetsov[1], Lynne R. Wilkens[1], Kami K. White[1], Marie Chong[1], Arielle Buyum[2], Grazyna Badowski[3], Rachael T. Leon Guerrero[4], Rachel Novotny[5]

[1]University of Hawaii Cancer Center, Honolulu, HI

[2]AB Consulting, LLC, Saipan, Northern Mariana Islands

[3]College of Natural & Applied Sciences, University of Guam, Mangilao, Guam

[4]Office of Research & Sponsored Programs, University of Guam, Mangilao, Guam

[5]College of Tropical Agriculture and Human Resources, University of Hawaii at Manoa, Honolulu, HI

Corresponding Author: Yurii B. Shvetsov, PhD, University of Hawaii Cancer Center, 701 Ilalo St, Honolulu HI 96813; tel: 808-564-5825; fax: 808-586-2982; email: yshvetso@cc.hawaii.edu.

Running Title: Prediction of breast cancer risk for Mariana Island women

Keywords: cancer: breast; cancer epidemiology; modeling

Financial Support: This work was supported by the U.S. National Cancer Institute, Comprehensive Partnerships to Reduce Cancer Health Disparities grants U54-CA143727 and U54-CA-143738, and by the U.S. National Cancer Institute grant R21-CA-220080.

Conflict of Interest: The authors declare no potential conflicts of interest.

Word count: 3218

Tables: 2

Figures: 3

ABSTRACT

Objectives: To develop a breast cancer risk prediction model for Chamorro and Filipino women of the Mariana Islands and compare its performance to that of the Breast Cancer Risk Assessment Tool (BCRAT).

Design: Case control study.

Setting: Clinics/facilities and other community-based settings on Guam and Saipan (Northern Mariana Islands).

Participants: 245 women (87 breast cancer cases and 148 controls) of Chamorro or Filipino ethnicity, age 25-80 years, with no prior history of cancer (other than skin cancer), residing on Guam or Saipan for at least 5 years.

Primary and secondary outcome measures: breast cancer risk models were constructed using combinations of exposures previously identified to affect breast cancer risk in this population, population breast cancer incidence rates and all-cause mortality rates for Guam.

Results: Models utilizing ethnic-specific relative risks performed better than those with relative risks estimated from all women. The model with the best performance among both ethnicities (the BRISK model; AUC: 0.66 and 0.65 among Chamorros and Filipinos, respectively) included age at first live birth and waist circumference. The 10-year breast cancer risk predicted by the BRISK model was 1.36% for Chamorros and 0.93% for Filipinos. Performance of the BCRAT was modest among both Chamorros (AUC: 0.60) and Filipinos (AUC: 0.55), possibly due to incomplete information on BCRAT risk factors.

Conclusions: The ability to develop breast cancer risk models for Mariana Islands women is constrained by the small population size and limited availability of health services and data.

Nonetheless, we have demonstrated that breast cancer risk prediction models with adequate discriminatory performance can be built for small populations such as in the Mariana Islands. Anthropometry, in particular waist circumference, was important for estimating breast cancer risk in this population.

ARTICLE SUMMARY

Strengths and limitations of the study

- The small sample size of this study is a direct consequence of the small population size.
- Our model construction method is designed to overcome the challenge of small population size.
- The final breast cancer risk model performed reasonably well.
- This is the first and only breast cancer risk prediction model for Chamorro and Filipino women of the Mariana Islands.

INTRODUCTION

Breast cancer is the most common cancer among women worldwide.[1] It is the second most common cause of cancer mortality among U.S. women [2] and has been the leading cause of cancer mortality among women on Guam over the last three decades.[3]

The Mariana Islands consist of two administrative units: Guam, a U.S. territory, and the Commonwealth of the Northern Mariana Islands (CNMI), which includes the islands of Saipan, Tinian, and Rota. The current population of Guam is ethnically mixed,[4] with 37% Chamorro, 26% Filipino, 12% other Pacific Islander, and 25% other ethnicity. CNMI is also diverse; its ethnic breakdown includes 24% Chamorro, 35% Filipino, 11% other Pacific Islander, and 30% other ethnicity.[5]

While the breast cancer incidence rate on Guam is lower than across the U.S., breast cancer mortality among some ethnicities on Guam, especially Chamorros, is higher than among U.S. women.[6] During 1998–2002 on Guam, the age-adjusted breast cancer incidence rate among Chamorro women was nearly twice as high as Filipino women and second only to White women (115.9, 60.7 and 148.6 per 100,000, respectively).[7] The age-adjusted incidence rate for U.S. women (not including data from the US affiliated Mariana Islands) during this time was 131. Chamorro

women also had the highest breast cancer mortality rate on Guam, at 32 per 100,000 women.[8] This contrasts with the overall U.S. mortality rate for that time period of 28 per 100,000.

The reasons for higher breast cancer mortality rates, and relatively high incidence rates, among Chamorro Pacific Islanders compared with other ethnic groups in the Mariana Islands are not well understood. The Breast Cancer Risk Model (BRISK) Project was conducted to improve understanding of the risk factors for breast cancer in this region.[9]

Estimation of a woman's breast cancer risk is an important tool used in primary breast cancer prevention efforts. One of the most widely used models for predicting breast cancer risk is the Gail model, developed for white women [10, 11] and subsequently extended to include other race/ethnicities such as African American and Asian American women.[12, 13] This extended model is available as NCI's Breast Cancer Risk Assessment Tool (BCRAT).[14] Although BCRAT includes Filipinos as one of the Asian American ethnicities, it is built from the Filipino population in SEER 9 registries,[15] whose age-specific breast cancer incidence rates differ from those for Filipinos on Guam, a US territory (Figure 1). A similar situation exists for Pacific Islanders, where only rates for Native Hawaiians are present in BCRAT. Additionally, BCRAT uses the same risk factors and relative risk estimates for all Asian American ethnicities; however, different breast cancer risk models are needed for adequate risk estimation for women of diverse racial/ethnic backgrounds,[16] and while some of the established risk factors are associated with breast cancer risk in the Mariana Islands women, others are not.[9] Due to these considerations, the utility of the BCRAT model for the Mariana Islands women is unknown.

In the present report, we evaluate performance of the BCRAT model and its modified version among Chamorro and Filipino participants in the BRISK study. In so doing, we propose a method of risk model development for small populations which we use here for the development and internal validation of a new breast cancer risk model for Chamorro and Filipino women of the Mariana Islands.

METHODS

BRISK study design and population

BRISK is a retrospective case-control study of mostly Asian and Pacific Islander women living on the Mariana Islands of Guam and Saipan. The study was a collaboration between the University of Guam and the University of Hawaii Cancer Center and was approved by the Institutional Review Boards at both institutions.

A detailed description of the study design and recruitment is provided elsewhere.[9, 17] Briefly, breast cancer cases and controls were recruited between 2010 and 2013. Breast cancer cases were identified through the Guam Cancer Registry, CNMI Department of Public Health, and health clinics on Guam. Controls were recruited in local clinics/facilities and other community-based settings on Guam and Saipan from among women with mammography screening and were frequency-matched to cases on age, ethnicity, and location (Saipan or Guam). Eligibility criteria for all participants were: (1) no prior history of cancer (other than skin cancer); (2) residence on Guam or Saipan for at least 5 years; (3) ability to provide consent for the study; and (4) age between 25 and 80 years. An additional eligibility criterion for cases was primary, invasive breast cancer newly diagnosed between 2009 and 2012.

During an interview, participants completed a detailed questionnaire including demographic, anthropometric, behavioral and lifestyle information; personal and family medical history; reproductive history; and acculturation based on a survey used in a multiethnic study.[18, 19] The reference date for the interview was the diagnosis date for cases and the interview date for controls. In addition, current waist circumference (WC), measured with an inelastic tape measure at the level of the umbilicus,[20] weight, height, and sitting height were measured by a trained anthropometrist. Body

mass index (BMI) was calculated as kg/m^2. Waist-height ratio (WHtR) was calculated as WC in cm divided by height in cm.

Of the 275 cases contacted, 38% agreed to participate, 21% were ineligible, and 41% refused due to scheduling conflicts, lack of transportation, family, psychological or cultural reasons, or off-island travel.[21] The corresponding percentages for controls were 74%, 20% and 6%. The study included 104 breast cancer cases (83 from Guam and 21 from CNMI) and 185 controls (140 from Guam and 45 from CNMI) between 27 and 80 years of age. A summary ethnicity variable was defined based on each participants' self-reported composition of her mother's and father's ethnicities. The present analysis was limited to participants with summary ethnicity of Chamorro and Filipino residing on Guam and Saipan (87 cases and 158 controls).

Patient and Public Involvement

Patients were not involved in the development of the research question, design of the study, recruitment and conduct of the study. However, the study provided funds to the CNMI Public Health mammography program to expand access and facilitate recruitment. The results were disseminated to study participants by public talks given at the University of Guam.

Breast cancer incidence and all-cause mortality rates

We obtained data from the Guam Cancer Registry (GCR) for all reportable female breast cancer diagnoses (n=576) on Guam for 2000-2009 (Supplementary Table S1).[17] Since data for CNMI were unavailable, Guam rates were also used to represent Saipan. Average annual age-specific incidence rates for female breast cancer were computed per ethnicity and 5-year age group, using interpolations between the U.S. 2000 and 2010 female census counts for Guam as denominators. All-cause mortality rates were obtained from the Guam Statistical Yearbook 2004.[22] Since 2004 was the only year these

rates were published, the rates for 2004 were used as a reasonable approximation for the 2000-2009 all-cause mortality rates.

Construction and selection of risk models

We assumed the general form of the Gail model,[10, 13, 23, 24] which projects absolute risk of breast cancer at a specified time interval using relative risk estimates for a set of risk factors, population breast cancer incidence rates and all-cause mortality rates. Risk factors considered for inclusion in the models were those identified in our previous report [9] as having a statistically significant ($p < 0.05$) association with breast cancer risk among Guam and Saipan women: age at first live birth (<20 or missing, 20-24, 25-29 or nulliparous, ≥30 y); BMI (<25, 25-29, ≥30); WHtR (≤0.54, 0.55-0.61 or missing, 0.62-0.67, >0.67); and WC (≤89, 90-99.5 or missing, >99.5 cm). Also considered for inclusion were the risk factors included in the original Gail model [10, 13] although they did not have a statistically significant association with breast cancer risk in our study: age at menarche (<12, 12-13, ≥14 y or missing); first-degree relatives with breast cancer (yes, no) and menopausal status (premenopausal, postmenopausal). As BMI, WHtR and WC were strongly correlated in our study, only one of these 3 factors was allowed to enter the model at a time. Following the approach of Gail et al.,[10] for each risk factor, missing values were grouped with the category showing the closest risk of breast cancer to participants with missing values, according to minimally adjusted logistic models. We constructed and evaluated models that included every combination of the above 7 risk factors as main effects (a total of 127 models). For each such combination, the entire dataset was used to estimate odds ratios (ORs) for the included risk factors using multivariable unconditional logistic regression, with adjustment for study participants' age, among both ethnicities combined and separately for Chamorros and Filipinos. Model-based adjusted attributable risk (AR) corresponding to these risk factors was then computed.[25] The Hosmer-Lemeshow statistic was computed to assess model fit. A risk model was

constructed using the OR and AR estimates from the logistic model. To assess model performance, a bootstrap validation method was utilized, whereby a validation subset was randomly selected, containing 50% of breast cancer cases (n=42) and two age and ethnicity-matched controls per case. The model was applied to all participants in the validation subset to project the absolute risk of breast cancer for a five-year period preceding the study interview date, and the area under the ROC curve (AUC) statistic was computed.

This bootstrap validation step was performed 100 times for each model, and the median AUC was computed. The top performing BRISK model was selected based on the highest median AUC for each ethnicity.

Evaluation of model performance

The final BRISK model was examined for its calibration and discrimination. The median AUC across bootstrap validation steps and its 95% confidence interval were taken as the measure of discriminatory performance of the model. Calibration of the model was assessed by examining the case/control distribution within quintiles of predicted 5- and 10-year absolute risk across the entire sample. The mean predicted risk of breast cancer was also computed for each quintile. Performance was compared with that of BCRAT.[13] As Native Hawaiians are the only Pacific Islander ethnicity represented in BCRAT and are the closest to Chamorros in terms of culture and lifestyle, we used Native Hawaiian incidence and mortality rates when applying BCRAT to Chamorro women. Due to a lack of breast biopsy information in our sample, all women were assumed to have had no breast biopsies, the default value in BCRAT.

Additionally, to examine whether calibrating the BCRAT model to the Guam breast cancer incidence rates would improve its performance, we modified the BCRAT model by replacing incidence and mortality rates with those for Filipino and Chamorro women on Guam, while retaining risk factors

and their relative risk estimates specified in the BCRAT; this modified model is referred to as BCRAT-G.

RESULTS

The demographic, lifestyle and reproductive characteristics of the study participants included in the present analysis (n=245) are summarized in Table 1. Briefly, the largest age group among both cases and controls was 50-59 years. One third of the participants (33%) were of Filipino ethnicity, the rest were Chamorros. The ethnic composition was similar among cases and controls by design, although the case to control ratio was somewhat higher among Filipino than Chamorro women (43% and 32% cases, respectively). Cases and controls had a similar proportion of women ever pregnant, pre-menopausal, parous, and having ever breastfed, but somewhat differed in BMI, WC, WHtR, alcohol consumption and smoking.

Table 1. Characteristics[a] of breast-cancer cases and controls among Chamorro and Filipino women of Mariana Islands in the BRISK study.

Characteristic, n (%)	Cases (n = 87)	Controls (n = 158)	P-value[b]
Age at reference[c], years (mean ± SD)	55.1 ± 10.8	53.8 ± 10.6	0.35
<40	7 (8.1)	12 (7.6)	0.92
40–49	22 (25.3)	47 (29.7)	
50–59	29 (33.3)	54 (34.2)	
60–69	19 (21.8)	31 (19.6)	
≥70	10 (11.5)	14 (8.9)	
Ethnicity			0.11
Chamorro	53 (60.9)	112 (70.9)	
Filipino	34 (39.1)	46 (29.1)	
Highest education level completed			0.99
High school diploma or less	40 (46.0)	73 (46.2)	
Some college	25 (28.7)	46 (29.1)	
College degree or more	22 (25.3)	39 (24.7)	
Age at menarche, years[c]			0.39
<12	20 (23.0)	45 (28.9)	
12–13	35 (40.2)	66 (42.3)	
≥14	32 (36.8)	45 (28.9)	

Ever been pregnant	78 (89.7)	145 (91.8)	0.58
Total number of pregnancies			0.67
0	9 (10.3)	13 (8.2)	
1–2	26 (29.9)	39 (24.7)	
3–4	31 (35.6)	59 (37.3)	
5 or more	21 (24.1)	47 (29.8)	
Number of live births			0.17
Nulliparous	10 (11.5)	17 (10.8)	
1–2	36 (41.4)	45 (28.5)	
3–4	25 (28.7)	63 (39.9)	
5 or more	16 (18.4)	33 (20.9)	
Age at first live birth, years, parous women only (mean ± SD)[c]	25.0 ± 5.5	22.9 ± 5.2	0.006
<20	18 (23.4)	48 (34.5)	0.03
20–24	22 (28.6)	52 (37.4)	
25–29	25 (32.5)	26 (18.7)	
≥30	12 (15.6)	13 (9.4)	
Ever breastfed, parous women only			0.83
No	24 (31.2)	42 (29.8)	
Yes	53 (68.8)	99 (70.2)	
Number of first-degree relatives with breast cancer			0.39
0	76 (87.4)	132 (83.5)	
1	8 (9.2)	23 (14.6)	
2	3 (3.4)	3 (1.9)	
Hormone use[c]			0.35
Never used estrogen or progesterone	77 (90.6)	133 (84.7)	
Yes, previously	8 (9.4)	21 (13.4)	
Yes, currently	0	3 (1.9)	
Menopausal status			0.21
Premenopausal	25 (28.7)	48 (30.4)	
Perimenopausal	4 (4.6)	17 (10.8)	
Postmenopausal	58 (66.7)	93 (58.9)	
Body mass Index, kg/m^2 (mean ± SD)	29.8 ± 7.0	30.3 ± 7.4	0.55
<18	0	0	0.10
18–24.9	18 (20.7)	44 (27.9)	
25–29.9	35 (40.2)	49 (31.0)	
≥30	34 (39.1)	65 (41.1)	
Waist Circumference, cm (mean ± SD)	97.2 ± 14.6	94.5 ± 14.9	0.19
Tertile 1 (≤89)[d]	24 (30.0)	53 (36.3)	0.54
Tertile 2 (89.1–99.5)	28 (35.0)	51 (34.9)	
Tertile 3 (>99.5)	28 (35.0)	42 (28.8)	
Waist/Height Ratio (mean ± SD)[e]	0.63 ± 0.09	0.61 ± 0.10	0.31
Quartile 1 (≤0.54)[d]	13 (16.3)	35 (24.0)	0.58
Quartile 2 (0.55–0.62)	27 (33.8)	43 (29.5)	
Quartile 3 (0.62–0.67)	20 (25.0)	32 (21.9)	
Quartile 4 (>0.67)	20 (25.0)	36 (24.7)	
Alcohol intake, drinks/week[e]			0.04
None	48 (76.2)	87 (61.7)	
Any alcohol reported	15 (23.8)	54 (38.3)	
Smoked daily for > 6 months[e]			0.05
No	61 (70.9)	92 (58.2)	

Yes	25 (29.1)	66 (41.8)

[a]Percentage is based on non-missing data and may not add up to 100 due to rounding.

[b]P-values based on chi-square test for categorical characteristics and t-test for continuous characteristics.

[c]Reference date was defined as diagnosis date for cases, interview date for controls.

[d]Quartiles and tertiles are based on the distribution among both cases and controls.

[e]Missing values were excluded: 2 controls for age at menarche, 2 controls for age at first live births, 2 cases and 1 control for hormone use, 7 cases and 12 controls for waist/height ratio, 24 cases and 17 controls for alcohol intake, 1 case for smoked daily for >6 months.

The composition of the top BRISK model and its performance are summarized in Table 2. The model included separate relative risk estimates among Chamorros and Filipinos for the included risk factors: age at menarche, age at first live birth and the number of first-degree relatives with breast cancer for both ethnicities, and additionally WC for Filipino women. The AUCs among Chamorros and Filipinos, respectively, were 0.64 and 0.67, based on the median across 100 validation runs.

The BRISK model classified more cases than controls into the highest risk stratum and more controls than cases into the lowest risk stratum among both ethnicities (Figures 2, 3), which indicates a good performance in terms of case/control distribution. Using case and control data, the BRISK model predicted a median 10-year absolute risk of breast cancer to be 1.28% for Chamorro women and 0.89% for Filipino women.

Table 2. Performance of the BRISK model and BCRAT among Mariana Island women in the BRISK study.

	BRISK[1,2]		BCRAT[3]		BCRAT-G[4]	
	Chamorros	Filipinos	Chamorros	Filipinos	Chamorros	Filipinos
Risk factors included / odds ratios:						
Age at menarche	1.134	1.710	1.078		1.078	
Age at first live birth	1.790	0.906	1.318		1.318	
Waist circumference	---	1.969	---		---	
Number of relatives with breast cancer	0.963	0.607	2.207		2.207	

Number of biopsies[5]	---	---	1.738		1.738	
Hosmer-Lemeshow statistic p-value[6]	0.52	0.86				
AUC (95% CI)[7]	0.64 (0.63 - 0.65)	0.67 (0.65 - 0.68)	0.60 (0.50 - 0.69)	0.55 (0.40 - 0.70)	0.59 (0.49 - 0.69)	0.51 (0.36 - 0.66)
Difference (% risk) in the median estimated risk between cases and controls[7]	0.33 (0.27-0.38)	0.31 (0.28-0.36)	0.18	0.13	0.06	0.00

[1]Highest AUC among Chamorros and Filipinos.

[2]Odds ratios for included risk factors are estimated in BRISK separately for Chamorros and Filipinos.

[3]BCRAT absolute risk estimates, selecting the Native Hawaiian BCRAT model for Chamorros.

[4]BCRAT absolute risk estimates, selecting the Native Hawaiian BCRAT model for Chamorros, with substitution of the breast cancer incidence and mortality rates by Guam rates.

[5]Number of biopsies was not available in our study and therefore was assigned the default value in the models.

[6]Computed using the underlying logistic regression model.

[7]Estimated as the median from 100 bootstrap validation datasets (30% data) for the BRISK model. Estimated using all data for BCRAT and BCRAT-G.

OR: odds ratio. AUC: area under the receiver operating characteristic curve. CI: confidence interval.

The unmodified BCRAT and the modified BCRAT-G model exhibited similar performance among Chamorros (AUC: 0.60 and 0.59, respectively) while BCRAT performed non-significantly better than BCRAT-G among Filipinos (AUC: 0.55 and 0.51, respectively; Table 2). Both models performed better among Chamorros than among Filipinos. Both BCRAT and BCRAT-G classified more controls than cases into the lower risk stratum among Filipinos, but not among Chamorros (Figures 2, 3). Both BCRAT and BCRAT-G classified more cases than controls into the higher risk stratum among both Chamorros and Filipinos.

DISCUSSION

To our knowledge, this is the first study that tested existing, as well as developed new, breast cancer risk models in a small, isolated population such as the Mariana Islands and in Pacific Islander populations other than Native Hawaiians. Developing or validating cancer risk models for populations such as Mariana Islands is challenging. Due to its unique ethnic composition and lifestyle, this population may be subject to unique risk factors not affecting other populations. The small population size places a natural restriction on the sample size of any epidemiologic study and reduces statistical power for potential model development. The population's geographic isolation results in the absence of sufficiently large comparable populations for external model validation.

A key challenge in our study was its small sample size, largely precipitated by the small size of the target population and newly emerging breast cancer registries. It is generally recommended that any new risk prediction model should include internal validation, either as bootstrap validation or utilizing training and validation subsets.[26, 27] As splitting a small dataset into training and validation parts would cause instability in the relative risk estimates and consequently in the resulting model, we have implemented a bootstrap validation procedure and used the entire dataset for parameter estimation. Our method produced a model that performed reasonably well, with AUC of 0.64-0.67 comparable to the AUC range of 0.53-0.68 for other published models.[28, 29] We also found that performance of the BCRAT model was modest among Chamorro and Filipino women in our study, with AUCs not exceeding 0.60. The poor performance of BCRAT-G indicates that replacing population incidence and mortality curves with those from the target population did not improve model performance.

There are several possible reasons that could explain the observed differences in model performance. First, in addition to the established risk factors in the Gail model, only the risk factors that exhibited significant associations with breast cancer risk in BRISK were considered for inclusion in the development of the model. Including risk factors not significantly associated with the outcome may cause model overfitting,[30] which may in turn bias the predicted absolute risk. In our previous

report,[9] no significant association between breast cancer risk and a number of known risk factors was found, but significant effects of several anthropometric factors such as BMI, WC and WHtR were observed on the risk of breast cancer. This may indicate a unique risk profile for this population or minimal variation in the known risk factors.

The BRISK model utilizes separate relative risk estimates by ethnicity with an additional risk factor (WC) for Filipinos; the model utilizing joint estimates did not perform as well. This indicates that Chamorros and Filipinos have different breast cancer risk profiles, which should be taken into account in risk prediction models. The BRISK model included anthropometrics in the form of WC, which reinforces the need to consider anthropometric measures in breast cancer risk models. Body size is dramatically different among the Asian and Pacific Islander residents in the Mariana Islands, with Filipino women generally having smaller body size than Chamorro women.[31, 32] BMI and central obesity have been found to be associated with higher breast cancer risk among Asian women,[33-35] and studies have demonstrated that the addition of body size variables improves prediction of breast cancer risk.[36] The inclusion of WC for Filipinos only may have to do with the issue of differing body sizes and excess overweight/obesity rates among Chamorros, thus diminishing the predictive value of body size for breast cancer in this ethnic group.

The BRISK model included 3-4 risk factors out of seven considered for inclusion. It has been suggested that the complexity threshold for a risk prediction model is 20 cases per model parameter.[26, 30] Exceeding this threshold in terms of the number of model parameters increases the danger of overfitting. In our study, with 87 breast cancer cases, the optimal number of model parameters is <5, which is evidenced in the final model. Applying a similar method of model selection and validation to a larger dataset may have resulted in a model with more parameters.

A recent focus of the breast cancer risk model improvement efforts has been examination of modifiable risk factors and their impact on predicted breast cancer risk.[37] The BRISK model

includes WC, a modifiable factor. This opens the possibility of the model being used as a supplemental health assessment tool in health behavior interventions, providing additional motivation for adoption of a healthier lifestyle that could decrease WC.

Limitations of our study include the small sample size as noted above, which may have prevented us from detecting important risk factors and, combined with limited response rate, may limit generalizability of findings. The failure to detect some expected associations (and thus to include the corresponding risk factors in the models) may also be due to a small sample size and lack of variability of some exposures in the study sample. The Guam breast cancer incidence rates covered a 10-year period and, thus, can be deemed reliable; however the all-cause mortality rates in our study are based on one year and thus may not be sufficiently stable. No CNMI breast cancer incidence or mortality rates were available and had to be approximated by the Guam rates. The information on risk exposures was limited; in particular, performance of the BCRAT model could have been affected by the lack of information on breast biopsies in our study.

Because BRISK was a case-control study, we were unable to assess model calibration to population incidence rates, although we examined the internal calibration of the model. We note, however, that the AUC-based comparison of models is robust to mis-calibration [24] and thus is a valid method in our study. We were also unable to perform external validation of the BRISK model, which is challenging given the unique nature and small size of this population, and remains a topic for future studies. Finally, AUCs based on the same dataset used for model construction may be overly optimistic.[30] We used the bootstrap validation method to minimize the optimism bias, although some of it may still persist. Despite these limitations, our model construction method has produced a reasonably well performing breast cancer risk model for Chamorro and Filipino women of the Mariana Islands, and the first and only model for this population.

CONCLUSIONS

We have demonstrated that breast cancer risk prediction models with adequate discriminatory performance can be built for small populations such as the Mariana Islands. The proposed model has the potential of being useful in primary breast cancer prevention in the Mariana Islands, but needs further refinement on larger samples of women and external validation on comparable Pacific Island populations.

ACKNOWLEDGEMENTS

We thank Michelle Blas-Laguana, Ashley Yamanaka, and Frances Santos-Hofschneider for recruiting and interviewing the BRISK Project participants, and the staff of Guam Radiology Consultants, FHP Clinic, and Guam Seventh-Day Adventist Clinic for their assistance in the recruitment of participants. We also thank the participants on Guam and Saipan who volunteered to take part in the BRISK study.

AUTHORS' CONTRIBUTIONS

Y.B.S. conducted the primary statistical analysis and had primary responsibility for the final manuscript; R.T.L.G. and R.N. led study concept and design; L.R.W. ensured integrity and accuracy of the study data; A.B. led data collection in Saipan; Y.B.S., L.R.W., K.K.W., M.C., G.B. contributed to statistical analysis; Y.B.S., L.R.W., K.K.W., R.T.L.G., R.N. interpreted the results and wrote the manuscript; Y.B.S., L.R.W., K.K.W., A.B., R.T.L.G., R.N. reviewed and approved the final manuscript.

FUNDING

This work was supported by the U.S. National Cancer Institute, Comprehensive Partnerships to Reduce Cancer Health Disparities grants U54-CA143727 and U54-CA-143738, and by the U.S. National Cancer Institute grant R21-CA-220080.

DATA SHARING

The datasets generated and analyzed during the current study are not publicly available because they contain protected health information. De-identified datasets are available from the senior author (R.N. at novotny@hawaii.edu) on reasonable request.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study has received ethical approval from the Institutional Review Boards at the University of Guam and the University of Hawaii. Written informed consent to participate was obtained from all study participants. All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

COMPETING INTERESTS

The authors declare that they have no competing interests to disclose.

FIGURE LEGEND

Figure 1. Cumulative incidence rates of invasive breast cancer in Guam and the U.S., 2000-2009.

Figure 2. Classification of breast cancer cases and controls into risk strata among Chamorro women in the BRISK study: the BRISK Model, BCRAT and BCRAT-G.

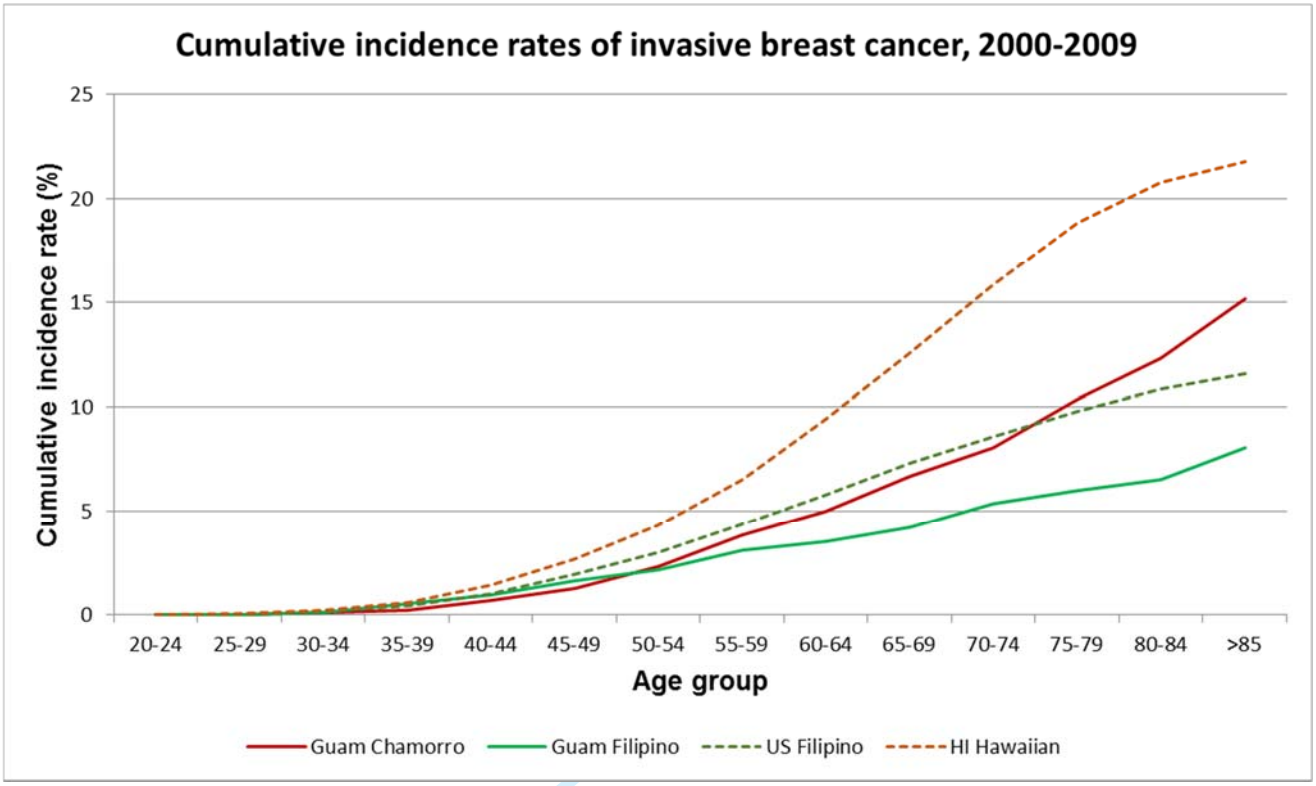
Figure 3. Classification of breast cancer cases and controls into risk strata among Filipino women in the BRISK study: the BRISK Model, BCRAT and BCRAT-G.

References

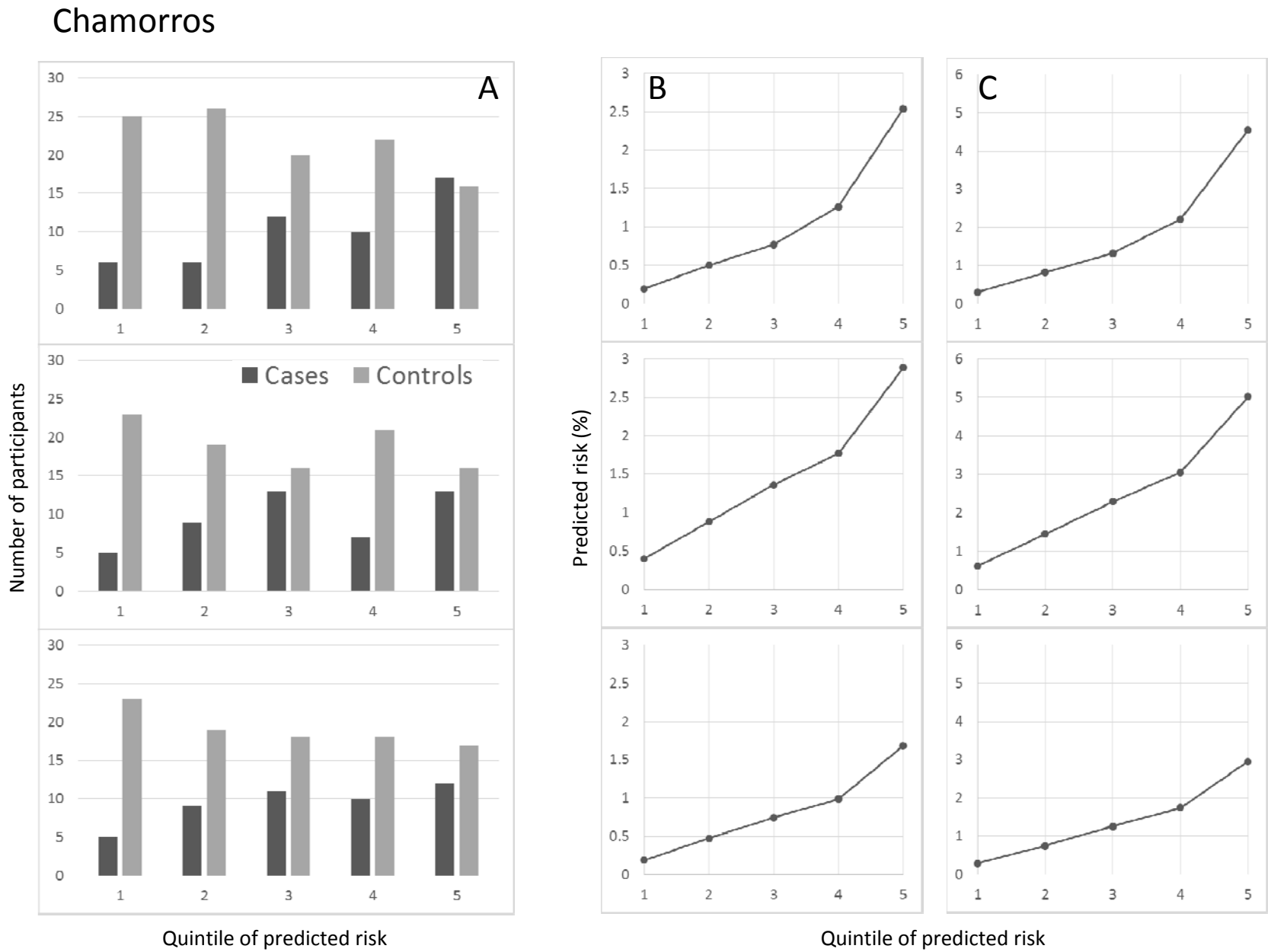
1. Ferlay J, Soerjomataram I, Dikshit R, *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359-86.
2. American Cancer Society. *Cancer facts & figures*. Atlanta, GA; 2018.
3. Haddock RL, Naval C. Cancer on Guam, especially among Micronesians. Pac Health Dialog. 2002;9:222–225.
4. Central Intelligence Agency. *The World Factbook*. https://www.cia.gov/library/publications/resources/the-world-factbook/geos/gq.html. Accessed November 19, 2018.
5. CNMI Department of Commerce. *Census demographics profile summary by district*. http://commerce.gov.mp/wp-content/uploads/2012/12/2010-Census-Demographics-Profile-Summary-by-District.pdf. Accessed November 19, 2018.
6. Haddock RL, Whippy HJ, Talon RJ, *et al.* Ethnic disparities in cancer incidence among residents of Guam. Asian Pac J Cancer Prev. 2009;10(1):57-62.
7. Guam Comprehensive Cancer Control Program. *Guam Cancer Facts & Figures 2008–2012*. Mangilao, GU; 2015.
8. Haddock RL, Talon RJ, Whippy HJ. Ethnic disparities in cancer mortality among residents of Guam. Asian Pac J Cancer Prev. 2006;7(3):411-4.
9. Leon Guerrero RT, Novotny R, Wilkens LR, *et al.* Risk factors for breast cancer in the breast cancer risk model study of Guam and Saipan. Cancer Epidemiol. 2017;50(Pt B):221-233.
10. Gail MH, Brinton LA, Byar DP, *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81(24):1879-86.
11. Costantino JP, Gail MH, Pee D, *et al.* Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst. 1999;91(18):1541-8.
12. Gail MH, Costantino JP, Pee D, *et al.* Projecting individualized absolute invasive breast cancer risk in African American women. J Natl Cancer Inst. 2007;99(23):1782-92.
13. Matsuno RK, Costantino JP, Ziegler RG, *et al.* Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. J Natl Cancer Inst. 2011;103(12):951-61.
14. *Breast Cancer Risk Assessment SAS Macro (Gail Model)*. http://dceg.cancer.gov/tools/risk-assessment/bcrasasmacro. Accessed September 5, 2017.
15. Surveillance E, and End Results (SEER) Program (www.seer.cancer.gov). SEER*Stat Database: Incidence - SEER Research Data, 9 Registries, Nov 2018 Sub (1975-2016) - Linked To County Attributes - Time Dependent (1990-2016) Income/Rurality, 1969-2017 Counties. In: National Cancer Institute D, Surveillance Research Program, (ed). November 2018 ed.
16. Bernstein L, Teal CR, Joslyn S, *et al.* Ethnicity-related variation in breast cancer risk factors. Cancer. 2003;97(1 Suppl):222-9.
17. Leon Guerrero RT, Badowski G, Yamanaka A, *et al.* University of Hawai'i Cancer Center connection: The vital role of cancer registries in the recruitment of an understudied minority population into a breast cancer study: Breast Cancer Risk Model for the Pacific. Hawaii J Med Public Health. 2014;73(10):335-40.
18. Goodman MT, Wu AH, Tung KH, *et al.* Association of dairy products, lactose, and calcium with the risk of ovarian cancer. Am J Epidemiol. 2002;156(2):148-57.
19. Tung KH, Goodman MT, Wu AH, *et al.* Reproductive factors and epithelial ovarian cancer risk by histologic type: a multiethnic case-control study. Am J Epidemiol. 2003;158(7):629-38.

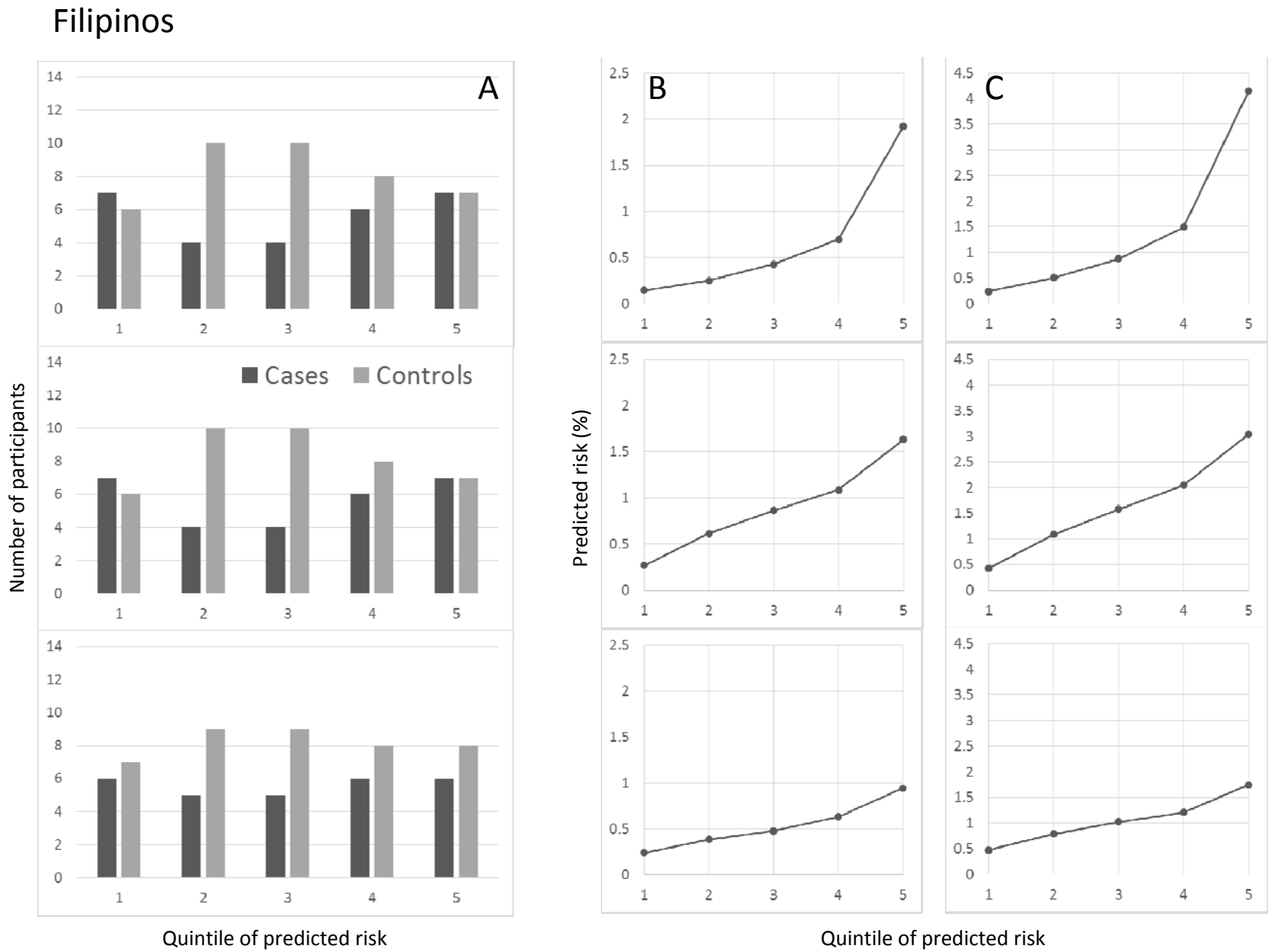
20. Lohman TG, Roche AF, Martorell R. *Anthropometric standardization reference manual*. Champaign, IL: Human Kinetics Books; 1988.
21. Leon Guerrero RT, Badowski G, Yamanaka A*, et al.* University of Hawai'i Cancer Center connection: The vital role of cancer registries in the recruitment of an understudied minority population into a breast cancer study: Breast Cancer Risk Model for the Pacific. Hawaii J Med Public Health. 2014;73(10):335-340.
22. Guam Bureau of Statistics and Plans. *Guam Statistical Yearbook*. Mangilao, GU; 2004.
23. Gail MH, Costantino JP. Validating and improving models for projecting the absolute risk of breast cancer. J Natl Cancer Inst. 2001;93(5):334-5.
24. Pfeiffer RM, Gail MH. *Absolute Risk: Methods and Applications in Clinical Management and Public Health*. Boca Raton, FL: CRC Press; 2017.
25. Bruzzi P, Green SB, Byar DP*, et al.* Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol. 1985;122(5):904-14.
26. Dupont WD, Blume JD, Smith JR. Building and Validating Complex Models of Breast Cancer Risk. JAMA Oncol. 2016;2(10):1271-1272.
27. Collins GS, Reitsma JB, Altman DG*, et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Ann Intern Med. 2015;162(10):735-6.
28. Anothaisintawee T, Teerawattananon Y, Wiratkapun C*, et al.* Risk prediction models of breast cancer: a systematic review of model performances. Breast Cancer Res Treat. 2012;133(1):1-10.
29. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat. 2012;132(2):365-77.
30. Harrell FE. *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
31. LeonGuerrero RT, Murphy SP, Novotny R*, et al.* Diet and obesity among Chamorro and Filipino Adults on Guam. Asia Pacific Journal of Clinical Nutrition. 2008;17(2):216-222.
32. Pinhey TK, Heathcote GM, Rarick J. The Influence of Obesity on the Self-Reported Health Status of Chamorros and other Residents of Guam. Asian Am Pac Isl J Health. 1994;2(3):195-211.
33. Bandera EV, Maskarinec G, Romieu I*, et al.* Racial and Ethnic Disparities in the Impact of Obesity on Breast Cancer Risk and Survival: A Global Perspective. Adv Nutr. 2015;6(6):803-19.
34. Nagrani R, Mhatre S, Rajaraman P*, et al.* Central obesity increases risk of breast cancer irrespective of menopausal and hormonal receptor status in women of South Asian Ethnicity. Eur J Cancer. 2016;66:153-61.
35. White KK, Park SY, Kolonel LN*, et al.* Body size and breast cancer risk: the Multiethnic Cohort. Int J Cancer. 2012;131(5):E705-16.
36. Chen J, Pee D, Ayyagari R*, et al.* Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. J Natl Cancer Inst. 2006;98(17):1215-26.
37. Maas P, Barrdahl M, Joshi AD*, et al.* Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA Oncol. 2016;2(10):1295-1302.

Figure 1. Cumulative incidence rates of invasive breast cancer in Guam and the U.S., 2000-2009.



Sources: (1) Guam Cancer Registry; (2) Hawaii Tumor Registry; (3) Surveillance, Epidemiology and End Results (SEER) 18-registry data.





Supplementary Table S1. Invasive breast cancer incidence rates in Guam women, 2000-2009 averages.

Age groups	Rate per 100,000	
	Filipino	Chamorro
20-24	0.0000	0.0000
25-29	6.8880	0.0000
30-34	19.8226	25.5028
35-39	87.2228	24.0964
40-44	76.7575	90.3465
45-49	137.0822	120.2289
50-54	110.8134	202.0321
55-59	186.8822	309.1133
60-64	85.5405	235.3095
65-69	139.3241	337.4036
70-74	227.2385	271.7201
75-79	123.1227	464.6494
80-84	108.3348	390.1445
>85	307.4785	564.2822

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *case-control studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(a) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	6
		(*b*) For matched studies, give matching criteria and the number of controls per case	6-7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-8
Bias	9	Describe any efforts to address potential sources of bias	6
Study size	10	Explain how the study size was arrived at	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-10
		(*b*) Describe any methods used to examine subgroups and interactions	8-10
		(*c*) Explain how missing data were addressed	8
		(*d*) If applicable, explain how matching of cases and controls was addressed	8
		(*e*) Describe any sensitivity analyses	9-10
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	7
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10-12
		(b) Indicate number of participants with missing data for each variable of interest	12

Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure	10-12
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	12-13
		(*b*) Report category boundaries when continuous variables were categorized	10-11
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	12
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12-13
Discussion			
Key results	18	Summarise key results with reference to study objectives	14
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	16
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	16
Generalisability	21	Discuss the generalisability (external validity) of the study results	14, 16
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17-18

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Prediction of breast cancer risk among women of the Mariana Islands: the BRISK model

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-061205.R1
Article Type:	Original research
Date Submitted by the Author:	01-Oct-2022
Complete List of Authors:	Shvetsov, Yurii; University of Hawai'i at Mānoa, Cancer Center Wilkens, Lynne ; University of Hawai'i at Mānoa, Cancer Center White, Kami; University of Hawai'i at Mānoa, Cancer Center Chong, Marie; University of Hawai'i at Mānoa, Cancer Center Buyum, Arielle; AB Consulting, LLC Badowski, Grazyna; University of Guam, College of Natural and Applied Sciences Leon Guerrero, Rachael; University of Guam, College of Natural and Applied Sciences Novotny, Rachel; University of Hawai'i at Manoa, College of Tropical Agriculture and Human Resources
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Oncology, Research methods
Keywords:	Breast tumours < ONCOLOGY, Epidemiology < ONCOLOGY, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Prediction of breast cancer risk among women of the Mariana Islands: the BRISK model

Yurii B. Shvetsov[1], Lynne R. Wilkens[1], Kami K. White[1], Marie Chong[1], Arielle Buyum[2], Grazyna Badowski[3], Rachael T. Leon Guerrero[4], Rachel Novotny[5]

[1]University of Hawaii Cancer Center, Honolulu, HI

[2]AB Consulting, LLC, Saipan, Northern Mariana Islands

[3]College of Natural & Applied Sciences, University of Guam, Mangilao, Guam

[4]Office of Research & Sponsored Programs, University of Guam, Mangilao, Guam

[5]College of Tropical Agriculture and Human Resources, University of Hawaii at Manoa, Honolulu, HI

Corresponding Author: Yurii B. Shvetsov, PhD, University of Hawaii Cancer Center, 701 Ilalo St, Honolulu HI 96813; tel: 808-564-5825; fax: 808-586-2982; email: yshvetso@cc.hawaii.edu.

Running Title: Prediction of breast cancer risk for Mariana Island women

Keywords: cancer: breast; cancer epidemiology; modeling

Financial Support: This work was supported by the U.S. National Cancer Institute, Comprehensive Partnerships to Reduce Cancer Health Disparities grants U54-CA143727 and U54-CA-143738, and by the U.S. National Cancer Institute grant R21-CA-220080.

Conflict of Interest: The authors declare no potential conflicts of interest.

Word count: 3218

Tables: 2

Figures: 3

ABSTRACT

Objectives: To develop a breast cancer risk prediction model for Chamorro and Filipino women of the Mariana Islands and compare its performance to that of the Breast Cancer Risk Assessment Tool (BCRAT).

Design: Case control study.

Setting: Clinics/facilities and other community-based settings on Guam and Saipan (Northern Mariana Islands).

Participants: 245 women (87 breast cancer cases and 148 controls) of Chamorro or Filipino ethnicity, age 25-80 years, with no prior history of cancer (other than skin cancer), residing on Guam or Saipan for at least 5 years.

Primary and secondary outcome measures: breast cancer risk models were constructed using combinations of exposures previously identified to affect breast cancer risk in this population, population breast cancer incidence rates and all-cause mortality rates for Guam.

Results: Models utilizing ethnic-specific relative risks performed better than those with relative risks estimated from all women. The model with the best performance among both ethnicities (the BRISK model; AUC: 0.66 and 0.65 among Chamorros and Filipinos, respectively) included age at first live birth and waist circumference. The 10-year breast cancer risk predicted by the BRISK model was 1.36% for Chamorros and 0.93% for Filipinos. Performance of the BCRAT was modest among both Chamorros (AUC: 0.60) and Filipinos (AUC: 0.55), possibly due to incomplete information on BCRAT risk factors.

Conclusions: The ability to develop breast cancer risk models for Mariana Islands women is constrained by the small population size and limited availability of health services and data.

Nonetheless, we have demonstrated that breast cancer risk prediction models with adequate discriminatory performance can be built for small populations such as in the Mariana Islands. Anthropometry, in particular waist circumference, was important for estimating breast cancer risk in this population.

ARTICLE SUMMARY

Strengths and limitations of the study

- The small sample size of this study is a direct consequence of the small population size.
- Our model construction method is designed to overcome the challenge of small population size.
- Bootstrap validation was used to minimize optimism bias.
- Evaluation of model coefficients separately for Chamorro and Filipino women of the Mariana Islands accounted for possible differential effect of model predictors between these two ethnic groups.

INTRODUCTION

Breast cancer is the most common cancer among women worldwide.[1] It is the second most common cause of cancer mortality among U.S. women [2] and has been the leading cause of cancer mortality among women on Guam over the last three decades.[3]

The Mariana Islands consist of two administrative units: Guam, a U.S. territory, and the Commonwealth of the Northern Mariana Islands (CNMI), which includes the islands of Saipan, Tinian, and Rota. The current population of Guam is ethnically mixed,[4] with 37% Chamorro, 26% Filipino, 12% other Pacific Islander, and 25% other ethnicity. CNMI is also diverse; its ethnic breakdown includes 24% Chamorro, 35% Filipino, 11% other Pacific Islander, and 30% other ethnicity.[5]

While the breast cancer incidence rate on Guam is lower than across the U.S., breast cancer mortality among some ethnicities on Guam, especially Chamorros, is higher than among U.S. women.[6] During 1998–2002 on Guam, the age-adjusted breast cancer incidence rate among Chamorro women was nearly twice as high as Filipino women and second only to White women (115.9, 60.7 and 148.6 per 100,000, respectively).[7] The age-adjusted incidence rate for U.S. women (not including data from the US affiliated Mariana Islands) during this time was 131. Chamorro

women also had the highest breast cancer mortality rate on Guam, at 32 per 100,000 women.[8] This contrasts with the overall U.S. mortality rate for that time period of 28 per 100,000.

The reasons for higher breast cancer mortality rates, and relatively high incidence rates, among Chamorro Pacific Islanders compared with other ethnic groups in the Mariana Islands are not well understood. The Breast Cancer Risk Model (BRISK) Project was conducted to improve understanding of the risk factors for breast cancer in this region.[9]

Estimation of a woman's breast cancer risk is an important tool used for risk assessment and stratification in breast cancer screening and prevention efforts. One of the most widely used models for predicting breast cancer risk is the Gail model, developed for white women [10, 11] and subsequently extended to include other race/ethnicities such as African American and Asian American women.[12, 13] This extended model is available as NCI's Breast Cancer Risk Assessment Tool (BCRAT).[14] Although BCRAT includes Filipinos as one of the Asian American ethnicities, it is built from the Filipino population in SEER 9 registries,[15] whose age-specific breast cancer incidence rates differ from those for Filipinos on Guam, a US territory (Figure 1). A similar situation exists for Pacific Islanders, where only rates for Native Hawaiians are present in BCRAT. Additionally, BCRAT uses the same risk factors and relative risk estimates for all Asian American ethnicities; however, different breast cancer risk models are needed for adequate risk estimation for women of diverse racial/ethnic backgrounds,[16] and while some of the established risk factors are associated with breast cancer risk in the Mariana Islands women, others are not.[9] Due to these considerations, the utility of the BCRAT model for the Mariana Islands women is unknown.

In the present report, we evaluate performance of the BCRAT model and its modified version among Chamorro and Filipino participants in the BRISK study. In so doing, we propose a method of risk model development for small populations which we use here for the development and internal validation of a new breast cancer risk model for Chamorro and Filipino women of the Mariana Islands.

METHODS

BRISK study design and population

BRISK is a retrospective case-control study of mostly Asian and Pacific Islander women living on the Mariana Islands of Guam and Saipan. The study was a collaboration between the University of Guam and the University of Hawaii Cancer Center and was approved by the Institutional Review Boards at both institutions.

A detailed description of the study design and recruitment is provided elsewhere.[9, 17] Briefly, breast cancer cases and controls were recruited between 2010 and 2013. Breast cancer cases were identified through the Guam Cancer Registry, CNMI Department of Public Health, and health clinics on Guam. Controls were recruited in local clinics/facilities and other community-based settings on Guam and Saipan from among women with mammography screening and were frequency-matched to cases on age, ethnicity, and location (Saipan or Guam). Eligibility criteria for all participants were: (1) no prior history of cancer (other than skin cancer); (2) residence on Guam or Saipan for at least 5 years; (3) ability to provide consent for the study; and (4) age between 25 and 80 years. An additional eligibility criterion for cases was primary, invasive breast cancer newly diagnosed between 2009 and 2012.

During an interview, participants completed a detailed questionnaire including demographic, anthropometric, behavioral and lifestyle information; personal and family medical history; reproductive history; and acculturation based on a survey used in a multiethnic study.[18, 19] The reference date for the interview was the diagnosis date for cases and the interview date for controls. In addition, current waist circumference (WC), measured with an inelastic tape measure at the level of the umbilicus,[20] weight, height, and sitting height were measured by a trained anthropometrist. Body

mass index (BMI) was calculated as kg/m^2. Waist-height ratio (WHtR) was calculated as WC in cm divided by height in cm.

Of the 275 cases contacted, 38% agreed to participate, 21% were ineligible, and 41% refused due to scheduling conflicts, lack of transportation, family, psychological or cultural reasons, or off-island travel.[21] The corresponding percentages for controls were 74%, 20% and 6%. The study included 104 breast cancer cases (83 from Guam and 21 from CNMI) and 185 controls (140 from Guam and 45 from CNMI) between 27 and 80 years of age. A summary ethnicity variable was defined based on each participants' self-reported composition of her mother's and father's ethnicities. The present analysis was limited to participants with summary ethnicity of Chamorro and Filipino residing on Guam and Saipan (87 cases and 158 controls).

Patient and Public Involvement

Patients were not involved in the development of the research question, design of the study, recruitment and conduct of the study. However, the study provided funds to the CNMI Public Health mammography program to expand access and facilitate recruitment. The results were disseminated to study participants by public talks given at the University of Guam.

Breast cancer incidence and all-cause mortality rates

We obtained data from the Guam Cancer Registry (GCR) for all reportable female breast cancer diagnoses (n=576) on Guam for 2000-2009 (Supplementary Table S1).[17] Since data for CNMI were unavailable, Guam rates were also used to represent Saipan. Average annual age-specific incidence rates for female breast cancer were computed per ethnicity and 5-year age group, using interpolations between the U.S. 2000 and 2010 female census counts for Guam as denominators. All-cause mortality rates were obtained from the Guam Statistical Yearbook 2004.[22] Since 2004 was the only year these

rates were published, the rates for 2004 were used as a reasonable approximation for the 2000-2009 all-cause mortality rates.

Construction and selection of risk models

We assumed the general form of the Gail model,[10, 13, 23, 24] which projects absolute risk of breast cancer at a specified time interval using relative risk estimates for a set of risk factors, population breast cancer incidence rates and all-cause mortality rates. Risk factors considered for inclusion in the models were those identified in our previous report [9] as having a statistically significant ($p < 0.05$) association with breast cancer risk among Guam and Saipan women: age at first live birth (<20 or missing, 20-24, 25-29 or nulliparous, ≥30 y); BMI (<25, 25-29, ≥30); WHtR (≤0.54, 0.55-0.61 or missing, 0.62-0.67, >0.67); and WC (≤89, 90-99.5 or missing, >99.5 cm). Also considered for inclusion were the risk factors included in the original Gail model [10, 13] although they did not have a statistically significant association with breast cancer risk in our study: age at menarche (<12, 12-13, ≥14 y or missing); first-degree relatives with breast cancer (yes, no) and menopausal status (premenopausal, postmenopausal). As BMI, WHtR and WC were strongly correlated in our study, only one of these 3 factors was allowed to enter the model at a time. Following the approach of Gail et al.,[10] for each risk factor, missing values were grouped with the category showing the closest risk of breast cancer to participants with missing values, according to minimally adjusted logistic models. We constructed and evaluated models that included every combination of the above 7 risk factors as main effects (a total of 127 models). For each such combination, the entire dataset was used to estimate odds ratios (ORs) for the included risk factors using multivariable unconditional logistic regression, with adjustment for study participants' age, among both ethnicities combined and separately for Chamorros and Filipinos. Model-based adjusted attributable risk (AR) corresponding to these risk factors was then computed.[25] The Hosmer-Lemeshow statistic was computed to assess model fit. A risk model was

constructed using the OR and AR estimates from the logistic model. To assess model performance, a bootstrap validation method was utilized, whereby a validation subset was randomly selected, containing 50% of breast cancer cases (n=42) and two age and ethnicity-matched controls per case. The model was applied to all participants in the validation subset to project the absolute risk of breast cancer for a five-year period preceding the study interview date, and the area under the ROC curve (AUC) statistic was computed.

This bootstrap validation step was performed 100 times for each model, and the median AUC was computed. The top performing BRISK model was selected based on the highest median AUC for each ethnicity.

Evaluation of model performance

The final BRISK model was examined for its calibration and discrimination. The median AUC across bootstrap validation steps and its 95% confidence interval were taken as the measure of discriminatory performance of the model. Calibration of the model was assessed by examining the case/control distribution within quintiles of predicted 5- and 10-year absolute risk across the entire sample. The mean predicted risk of breast cancer was also computed for each quintile. Performance was compared with that of BCRAT.[13] As Native Hawaiians are the only Pacific Islander ethnicity represented in BCRAT and are the closest to Chamorros in terms of culture and lifestyle, we used Native Hawaiian incidence and mortality rates when applying BCRAT to Chamorro women. Due to a lack of breast biopsy information in our sample, all women were assumed to have had no breast biopsies, the default value in BCRAT.

Additionally, to examine whether calibrating the BCRAT model to the Guam breast cancer incidence rates would improve its performance, we modified the BCRAT model by replacing incidence and mortality rates with those for Filipino and Chamorro women on Guam, while retaining risk factors

and their relative risk estimates specified in the BCRAT; this modified model is referred to as BCRAT-G.

RESULTS

The demographic, lifestyle and reproductive characteristics of the study participants included in the present analysis (n=245) are summarized in Table 1. Briefly, the largest age group among both cases and controls was 50-59 years. One third of the participants (33%) were of Filipino ethnicity, the rest were Chamorros. The ethnic composition was similar among cases and controls by design, although the case to control ratio was somewhat higher among Filipino than Chamorro women (43% and 32% cases, respectively). Cases and controls had a similar proportion of women ever pregnant, pre-menopausal, parous, and having ever breastfed, but somewhat differed in BMI, WC, WHtR, alcohol consumption and smoking.

Table 1. Characteristics[a] of breast-cancer cases and controls among Chamorro and Filipino women of Mariana Islands in the BRISK study.

Characteristic, n (%)	Cases (n = 87)	Controls (n = 158)	P-value[b]
Age at reference[c], years (mean ± SD)	55.1 ± 10.8	53.8 ± 10.6	0.35
<40	7 (8.1)	12 (7.6)	0.92
40–49	22 (25.3)	47 (29.7)	
50–59	29 (33.3)	54 (34.2)	
60–69	19 (21.8)	31 (19.6)	
≥70	10 (11.5)	14 (8.9)	
Ethnicity			0.11
Chamorro	53 (60.9)	112 (70.9)	
Filipino	34 (39.1)	46 (29.1)	
Highest education level completed			0.99
High school diploma or less	40 (46.0)	73 (46.2)	
Some college	25 (28.7)	46 (29.1)	
College degree or more	22 (25.3)	39 (24.7)	
Age at menarche, years[c]			0.39
<12	20 (23.0)	45 (28.9)	
12–13	35 (40.2)	66 (42.3)	
≥14	32 (36.8)	45 (28.9)	

Ever been pregnant	78 (89.7)	145 (91.8)	0.58
Total number of pregnancies			0.67
0	9 (10.3)	13 (8.2)	
1–2	26 (29.9)	39 (24.7)	
3–4	31 (35.6)	59 (37.3)	
5 or more	21 (24.1)	47 (29.8)	
Number of live births			0.17
Nulliparous	10 (11.5)	17 (10.8)	
1–2	36 (41.4)	45 (28.5)	
3–4	25 (28.7)	63 (39.9)	
5 or more	16 (18.4)	33 (20.9)	
Age at first live birth, years, parous women only (mean ± SD)[c]	25.0 ± 5.5	22.9 ± 5.2	0.006
<20	18 (23.4)	48 (34.5)	0.03
20–24	22 (28.6)	52 (37.4)	
25–29	25 (32.5)	26 (18.7)	
≥30	12 (15.6)	13 (9.4)	
Ever breastfed, parous women only			0.83
No	24 (31.2)	42 (29.8)	
Yes	53 (68.8)	99 (70.2)	
Number of first-degree relatives with breast cancer			0.39
0	76 (87.4)	132 (83.5)	
1	8 (9.2)	23 (14.6)	
2	3 (3.4)	3 (1.9)	
Hormone use[c]			0.35
Never used estrogen or progesterone	77 (90.6)	133 (84.7)	
Yes, previously	8 (9.4)	21 (13.4)	
Yes, currently	0	3 (1.9)	
Menopausal status			0.21
Premenopausal	25 (28.7)	48 (30.4)	
Perimenopausal	4 (4.6)	17 (10.8)	
Postmenopausal	58 (66.7)	93 (58.9)	
Body mass Index, kg/m^2 (mean ± SD)	29.8 ± 7.0	30.3 ± 7.4	0.55
<18	0	0	0.10
18–24.9	18 (20.7)	44 (27.9)	
25–29.9	35 (40.2)	49 (31.0)	
≥30	34 (39.1)	65 (41.1)	
Waist Circumference, cm (mean ± SD)	97.2 ± 14.6	94.5 ± 14.9	0.19
Tertile 1 (≤89)[d]	24 (30.0)	53 (36.3)	0.54
Tertile 2 (89.1–99.5)	28 (35.0)	51 (34.9)	
Tertile 3 (>99.5)	28 (35.0)	42 (28.8)	
Waist/Height Ratio (mean ± SD)[e]	0.63 ± 0.09	0.61 ± 0.10	0.31
Quartile 1 (≤0.54)[d]	13 (16.3)	35 (24.0)	0.58
Quartile 2 (0.55–0.62)	27 (33.8)	43 (29.5)	
Quartile 3 (0.62–0.67)	20 (25.0)	32 (21.9)	
Quartile 4 (>0.67)	20 (25.0)	36 (24.7)	
Alcohol intake, drinks/week[e]			0.04
None	48 (76.2)	87 (61.7)	
Any alcohol reported	15 (23.8)	54 (38.3)	
Smoked daily for > 6 months[e]			0.05
No	61 (70.9)	92 (58.2)	

Yes	25 (29.1)	66 (41.8)

[a]Percentage is based on non-missing data and may not add up to 100 due to rounding.

[b]P-values based on chi-square test for categorical characteristics and t-test for continuous characteristics.

[c]Reference date was defined as diagnosis date for cases, interview date for controls.

[d]Quartiles and tertiles are based on the distribution among both cases and controls.

[e]Missing values were excluded: 2 controls for age at menarche, 2 controls for age at first live births, 2 cases and 1 control for hormone use, 7 cases and 12 controls for waist/height ratio, 24 cases and 17 controls for alcohol intake, 1 case for smoked daily for >6 months.

The composition of the top BRISK model and its performance are summarized in Table 2. The model included separate relative risk estimates among Chamorros and Filipinos for the included risk factors: age at menarche, age at first live birth and the number of first-degree relatives with breast cancer for both ethnicities, and additionally WC for Filipino women. The AUCs among Chamorros and Filipinos, respectively, were 0.64 and 0.67, based on the median across 100 validation runs.

The BRISK model classified more cases than controls into the highest risk stratum and more controls than cases into the lowest risk stratum among both ethnicities (Figures 2, 3), which indicates a good performance in terms of case/control distribution. Using case and control data, the BRISK model predicted a median 10-year absolute risk of breast cancer to be 1.28% for Chamorro women and 0.89% for Filipino women.

Table 2. Performance of the BRISK model and BCRAT among Mariana Island women in the BRISK study.

	BRISK[1,2]		BCRAT[3]		BCRAT-G[4]	
	Chamorros	Filipinos	Chamorros	Filipinos	Chamorros	Filipinos
Risk factors included / odds ratios:						
Age at menarche	1.13 (0.72-1.78)	1.71 (0.84-3.49)	1.078		1.078	
Age at first live birth	1.79 (1.25-2.56)	0.91 (0.58-1.41)	1.318		1.318	
Waist circumference	---	1.97 (1.19-3.25)	---		---	

Number of relatives with breast cancer	0.96 (0.39-2.36)	0.61 (0.09-4.07)	2.207		2.207	
Number of biopsies[5]	---	---	1.738		1.738	
Hosmer-Lemeshow statistic p-value[6]	0.52	0.86				
AUC (95% CI)[7]	0.64 (0.63 - 0.65)	0.67 (0.65 - 0.68)	0.60 (0.50 - 0.69)	0.55 (0.40 - 0.70)	0.59 (0.49 - 0.69)	0.51 (0.36 - 0.66)
Difference (% risk) in the median estimated risk between cases and controls[7]	0.33 (0.27-0.38)	0.31 (0.28-0.36)	0.18	0.13	0.06	0.00

[1]Highest AUC among Chamorros and Filipinos.

[2]Odds ratios for included risk factors are estimated in BRISK separately for Chamorros and Filipinos.

[3]BCRAT absolute risk estimates, selecting the Native Hawaiian BCRAT model for Chamorros.

[4]BCRAT absolute risk estimates, selecting the Native Hawaiian BCRAT model for Chamorros, with substitution of the breast cancer incidence and mortality rates by Guam rates.

[5]Number of biopsies was not available in our study and therefore was assigned the default value in the models.

[6]Computed using the underlying logistic regression model.

[7]Estimated as the median from 100 bootstrap validation datasets (30% data) for the BRISK model. Estimated using all data for BCRAT and BCRAT-G.

OR: odds ratio. AUC: area under the receiver operating characteristic curve. CI: confidence interval.

The unmodified BCRAT and the modified BCRAT-G model exhibited similar performance among Chamorros (AUC: 0.60 and 0.59, respectively) while BCRAT performed non-significantly better than BCRAT-G among Filipinos (AUC: 0.55 and 0.51, respectively; Table 2). Both models performed better among Chamorros than among Filipinos. Both BCRAT and BCRAT-G classified more controls than cases into the lower risk stratum among Filipinos, but not among Chamorros (Figures 2, 3). Both BCRAT and BCRAT-G classified more cases than controls into the higher risk stratum among both Chamorros and Filipinos.

DISCUSSION

To our knowledge, this is the first study that tested existing, as well as developed new, breast cancer risk models in a small, isolated population such as the Mariana Islands and in Pacific Islander populations other than Native Hawaiians. Developing or validating cancer risk models for populations such as Mariana Islands is challenging. Due to its unique ethnic composition and lifestyle, this population may be subject to unique risk factors not affecting other populations. The small population size places a natural restriction on the sample size of any epidemiologic study and reduces statistical power for potential model development. The population's geographic isolation results in the absence of sufficiently large comparable populations for external model validation.

A key challenge in our study was its small sample size, largely precipitated by the small size of the target population and newly emerging breast cancer registries. It is generally recommended that any new risk prediction model should include internal validation, either as bootstrap validation or utilizing training and validation subsets.[26, 27] As splitting a small dataset into training and validation parts would cause instability in the relative risk estimates and consequently in the resulting model, we have implemented a bootstrap validation procedure and used the entire dataset for parameter estimation. Our method produced a model that performed reasonably well, with AUC of 0.64-0.67 comparable to the AUC range of 0.53-0.68 for other published models.[28, 29] We also found that performance of the BCRAT model was modest among Chamorro and Filipino women in our study, with AUCs not exceeding 0.60. The poor performance of BCRAT-G indicates that replacing population incidence and mortality curves with those from the target population did not improve model performance.

There are several possible reasons that could explain the observed differences in model performance. First, in addition to the established risk factors in the Gail model, only the risk factors that exhibited significant associations with breast cancer risk in BRISK were considered for inclusion in the development of the model. Including risk factors not significantly associated with the outcome may cause model overfitting,[30] which may in turn bias the predicted absolute risk. In our previous

report,[9] no significant association between breast cancer risk and a number of known risk factors was found, but significant effects of several anthropometric factors such as BMI, WC and WHtR were observed on the risk of breast cancer. This may indicate a unique risk profile for this population or minimal variation in the known risk factors.

The BRISK model utilizes separate relative risk estimates by ethnicity with an additional risk factor (WC) for Filipinos; the model utilizing joint estimates did not perform as well. This indicates that Chamorros and Filipinos have different breast cancer risk profiles, which should be taken into account in risk prediction models. The BRISK model included anthropometrics in the form of WC, which reinforces the need to consider anthropometric measures in breast cancer risk models. Body size is dramatically different among the Asian and Pacific Islander residents in the Mariana Islands, with Filipino women generally having smaller body size than Chamorro women.[31, 32] BMI and central obesity have been found to be associated with higher breast cancer risk among Asian women,[33-35] and studies have demonstrated that the addition of body size variables improves prediction of breast cancer risk.[36] The inclusion of WC for Filipinos only may have to do with the issue of differing body sizes and excess overweight/obesity rates among Chamorros, thus diminishing the predictive value of body size for breast cancer in this ethnic group.

The BRISK model included 3-4 risk factors out of seven considered for inclusion. It has been suggested that the complexity threshold for a risk prediction model is 20 cases per model parameter.[26, 30] Exceeding this threshold in terms of the number of model parameters increases the danger of overfitting. In our study, with 87 breast cancer cases, the optimal number of model parameters is <5, which is evidenced in the final model. Applying a similar method of model selection and validation to a larger dataset may have resulted in a model with more parameters.

A recent focus of the breast cancer risk model improvement efforts has been examination of modifiable risk factors and their impact on predicted breast cancer risk.[37] The BRISK model

includes WC, a modifiable factor. This opens the possibility of the model being used as a supplemental health assessment tool in health behavior interventions, providing additional motivation for adoption of a healthier lifestyle that could decrease WC. As all predictors in the model can be collected from a patient questionnaire, the model can easily be implemented in most clinic settings including local clinics.

Limitations of our study include the small sample size as noted above, which may have prevented us from detecting important risk factors and, combined with limited response rate, may limit generalizability of findings. The failure to detect some expected associations (and thus to include the corresponding risk factors in the models) may also be due to a small sample size and lack of variability of some exposures in the study sample. The information on risk exposures was limited; in particular, performance of the BCRAT model could have been affected by the lack of information on breast biopsies in our study. The Guam breast cancer incidence rates covered a 10-year period and, thus, can be deemed reliable; however the all-cause mortality rates in our study are based on one year and thus may not be sufficiently stable. No CNMI breast cancer incidence or mortality rates were available and had to be approximated by the Guam rates. Although the ethnic composition of CNMI is similar to Guam, it is possible that, due to differing lifestyle and poorer access to healthcare among the CNMI population, the breast cancer incidence and overall mortality rates in CNMI differ from those on Guam. However, since the majority of study participants were from Guam, this potential difference in rates was unlikely to have a substantial impact on our results. Nonetheless, efforts are needed to collect and disseminate data on cancer incidence and mortality rates for CNMI, which would allow researchers to improve study results.

Because BRISK was a case-control study, we were unable to assess model calibration to population incidence rates, although we examined the internal calibration of the model. We note, however, that the AUC-based comparison of models is robust to mis-calibration [24] and thus is a valid

method in our study. We were also unable to perform external validation of the BRISK model, which is challenging given the unique nature and small size of this population, and remains a topic for future studies. Finally, AUCs based on the same dataset used for model construction may be overly optimistic.[30] We used the bootstrap validation method to minimize the optimism bias, although some of it may still persist. Despite these limitations, our model construction method has produced a reasonably well performing breast cancer risk model for Chamorro and Filipino women of the Mariana Islands, and the first and only model for this population.

CONCLUSIONS

We have demonstrated that breast cancer risk prediction models with adequate discriminatory performance can be built for small populations such as the Mariana Islands. The proposed model has the potential of being useful as a supplemental tool for risk assessment and stratification in breast cancer screening and prevention in the Mariana Islands, but needs further refinement on larger samples of women and external validation on comparable Pacific Island populations.

ACKNOWLEDGEMENTS

We thank Michelle Blas-Laguana, Ashley Yamanaka, and Frances Santos-Hofschneider for recruiting and interviewing the BRISK Project participants, and the staff of Guam Radiology Consultants, FHP Clinic, and Guam Seventh-Day Adventist Clinic for their assistance in the recruitment of participants. We also thank the participants on Guam and Saipan who volunteered to take part in the BRISK study.

AUTHORS' CONTRIBUTIONS

Y.B.S. conducted the primary statistical analysis and had primary responsibility for the final manuscript; R.T.L.G. and R.N. led study concept and design; L.R.W. ensured integrity and accuracy of

the study data; A.B. led data collection in Saipan; Y.B.S., L.R.W., K.K.W., M.C., G.B. contributed to statistical analysis; Y.B.S., L.R.W., K.K.W., R.T.L.G., R.N. interpreted the results and wrote the manuscript; Y.B.S., L.R.W., K.K.W., A.B., R.T.L.G., R.N. reviewed and approved the final manuscript.

FUNDING

This work was supported by the U.S. National Cancer Institute, Comprehensive Partnerships to Reduce Cancer Health Disparities grants U54-CA143727 and U54-CA-143738, and by the U.S. National Cancer Institute grant R21-CA-220080.

DATA SHARING

The datasets generated and analyzed during the current study are not publicly available because they contain protected health information. De-identified datasets are available from the senior author (R.N. at novotny@hawaii.edu) on reasonable request.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study has received ethical approval from the Institutional Review Boards at the University of Guam and the University of Hawaii. Written informed consent to participate was obtained from all study participants. All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

COMPETING INTERESTS

The authors declare that they have no competing interests to disclose.

FIGURE LEGEND

Figure 1. Cumulative incidence rates of invasive breast cancer in Guam and the U.S., 2000-2009.

Figure 2. Classification of breast cancer cases and controls into risk strata among Chamorro women in the BRISK study: the BRISK Model, BCRAT and BCRAT-G. A: Frequency of cases and controls by quintile of predicted risk. B: Mean predicted 5-year risk by quintile of predicted risk. C: Mean predicted 10-year risk by quintile of predicted risk.

Figure 3. Classification of breast cancer cases and controls into risk strata among Filipino women in the BRISK study: the BRISK Model, BCRAT and BCRAT-G. A: Frequency of cases and controls by quintile of predicted risk. B: Mean predicted 5-year risk by quintile of predicted risk. C: Mean predicted 10-year risk by quintile of predicted risk.

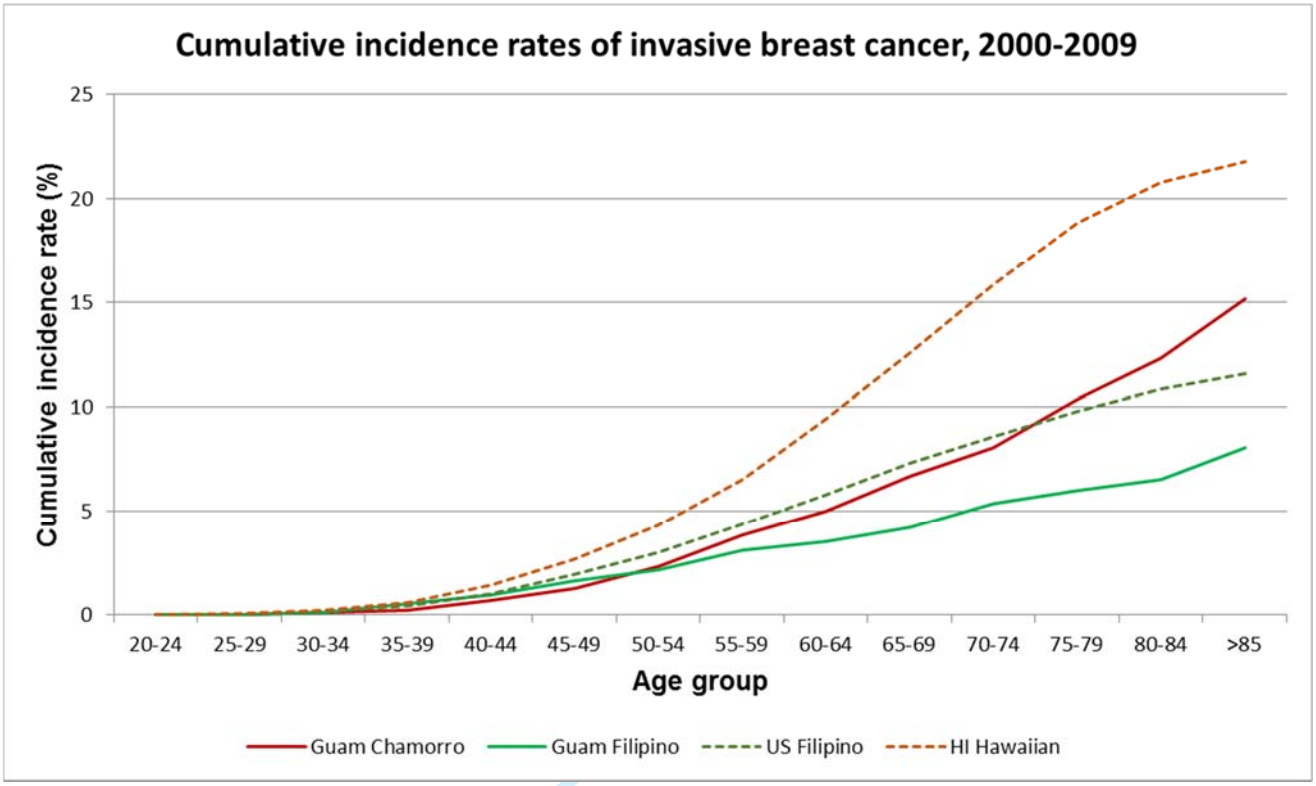
References

1. Ferlay J, Soerjomataram I, Dikshit R, *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359-86.
2. American Cancer Society. *Cancer facts & figures*. Atlanta, GA; 2018.
3. Haddock RL, Naval C. Cancer on Guam, especially among Micronesians. Pac Health Dialog. 2002;9:222–225.
4. Central Intelligence Agency. *The World Factbook*. https://www.cia.gov/library/publications/resources/the-world-factbook/geos/gq.html. Accessed November 19, 2018.
5. CNMI Department of Commerce. *Census demographics profile summary by district*. http://commerce.gov.mp/wp-content/uploads/2012/12/2010-Census-Demographics-Profile-Summary-by-District.pdf. Accessed November 19, 2018.
6. Haddock RL, Whippy HJ, Talon RJ, *et al.* Ethnic disparities in cancer incidence among residents of Guam. Asian Pac J Cancer Prev. 2009;10(1):57-62.
7. Guam Comprehensive Cancer Control Program. *Guam Cancer Facts & Figures 2008–2012*. Mangilao, GU; 2015.
8. Haddock RL, Talon RJ, Whippy HJ. Ethnic disparities in cancer mortality among residents of Guam. Asian Pac J Cancer Prev. 2006;7(3):411-4.
9. Leon Guerrero RT, Novotny R, Wilkens LR, *et al.* Risk factors for breast cancer in the breast cancer risk model study of Guam and Saipan. Cancer Epidemiol. 2017;50(Pt B):221-233.
10. Gail MH, Brinton LA, Byar DP, *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81(24):1879-86.
11. Costantino JP, Gail MH, Pee D, *et al.* Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst. 1999;91(18):1541-8.

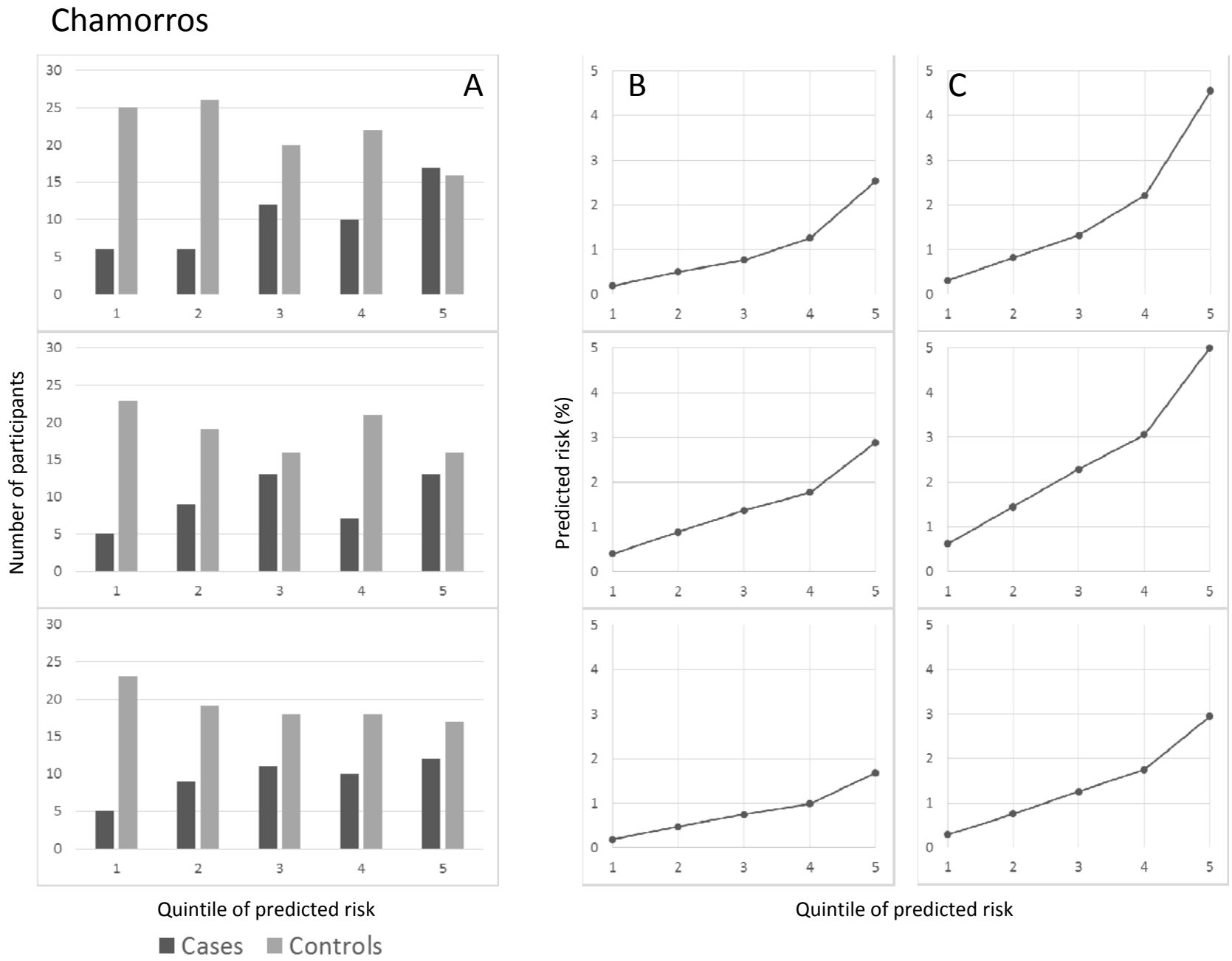
12. Gail MH, Costantino JP, Pee D, *et al.* Projecting individualized absolute invasive breast cancer risk in African American women. J Natl Cancer Inst. 2007;99(23):1782-92.
13. Matsuno RK, Costantino JP, Ziegler RG, *et al.* Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. J Natl Cancer Inst. 2011;103(12):951-61.
14. *Breast Cancer Risk Assessment SAS Macro (Gail Model)*. http://dceg.cancer.gov/tools/risk-assessment/bcrasasmacro. Accessed September 5, 2017.
15. Surveillance E, and End Results (SEER) Program (www.seer.cancer.gov). SEER*Stat Database: Incidence - SEER Research Data, 9 Registries, Nov 2018 Sub (1975-2016) - Linked To County Attributes - Time Dependent (1990-2016) Income/Rurality, 1969-2017 Counties. In: National Cancer Institute D, Surveillance Research Program, (ed). November 2018 ed.
16. Bernstein L, Teal CR, Joslyn S, *et al.* Ethnicity-related variation in breast cancer risk factors. Cancer. 2003;97(1 Suppl):222-9.
17. Leon Guerrero RT, Badowski G, Yamanaka A, *et al.* University of Hawai'i Cancer Center connection: The vital role of cancer registries in the recruitment of an understudied minority population into a breast cancer study: Breast Cancer Risk Model for the Pacific. Hawaii J Med Public Health. 2014;73(10):335-40.
18. Goodman MT, Wu AH, Tung KH, *et al.* Association of dairy products, lactose, and calcium with the risk of ovarian cancer. Am J Epidemiol. 2002;156(2):148-57.
19. Tung KH, Goodman MT, Wu AH, *et al.* Reproductive factors and epithelial ovarian cancer risk by histologic type: a multiethnic case-control study. Am J Epidemiol. 2003;158(7):629-38.
20. Lohman TG, Roche AF, Martorell R. *Anthropometric standardization reference manual*. Champaign, IL: Human Kinetics Books; 1988.
21. Leon Guerrero RT, Badowski G, Yamanaka A, *et al.* University of Hawai'i Cancer Center connection: The vital role of cancer registries in the recruitment of an understudied minority population into a breast cancer study: Breast Cancer Risk Model for the Pacific. Hawaii J Med Public Health. 2014;73(10):335-340.
22. Guam Bureau of Statistics and Plans. *Guam Statistical Yearbook*. Mangilao, GU; 2004.
23. Gail MH, Costantino JP. Validating and improving models for projecting the absolute risk of breast cancer. J Natl Cancer Inst. 2001;93(5):334-5.
24. Pfeiffer RM, Gail MH. *Absolute Risk: Methods and Applications in Clinical Management and Public Health*. Boca Raton, FL: CRC Press; 2017.
25. Bruzzi P, Green SB, Byar DP, *et al.* Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol. 1985;122(5):904-14.
26. Dupont WD, Blume JD, Smith JR. Building and Validating Complex Models of Breast Cancer Risk. JAMA Oncol. 2016;2(10):1271-1272.
27. Collins GS, Reitsma JB, Altman DG, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Ann Intern Med. 2015;162(10):735-6.
28. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, *et al.* Risk prediction models of breast cancer: a systematic review of model performances. Breast Cancer Res Treat. 2012;133(1):1-10.
29. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat. 2012;132(2):365-77.
30. Harrell FE. *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
31. LeonGuerrero RT, Murphy SP, Novotny R, *et al.* Diet and obesity among Chamorro and Filipino Adults on Guam. Asia Pacific Journal of Clinical Nutrition. 2008;17(2):216-222.
32. Pinhey TK, Heathcote GM, Rarick J. The Influence of Obesity on the Self-Reported Health Status of Chamorros and other Residents of Guam. Asian Am Pac Isl J Health. 1994;2(3):195-211.

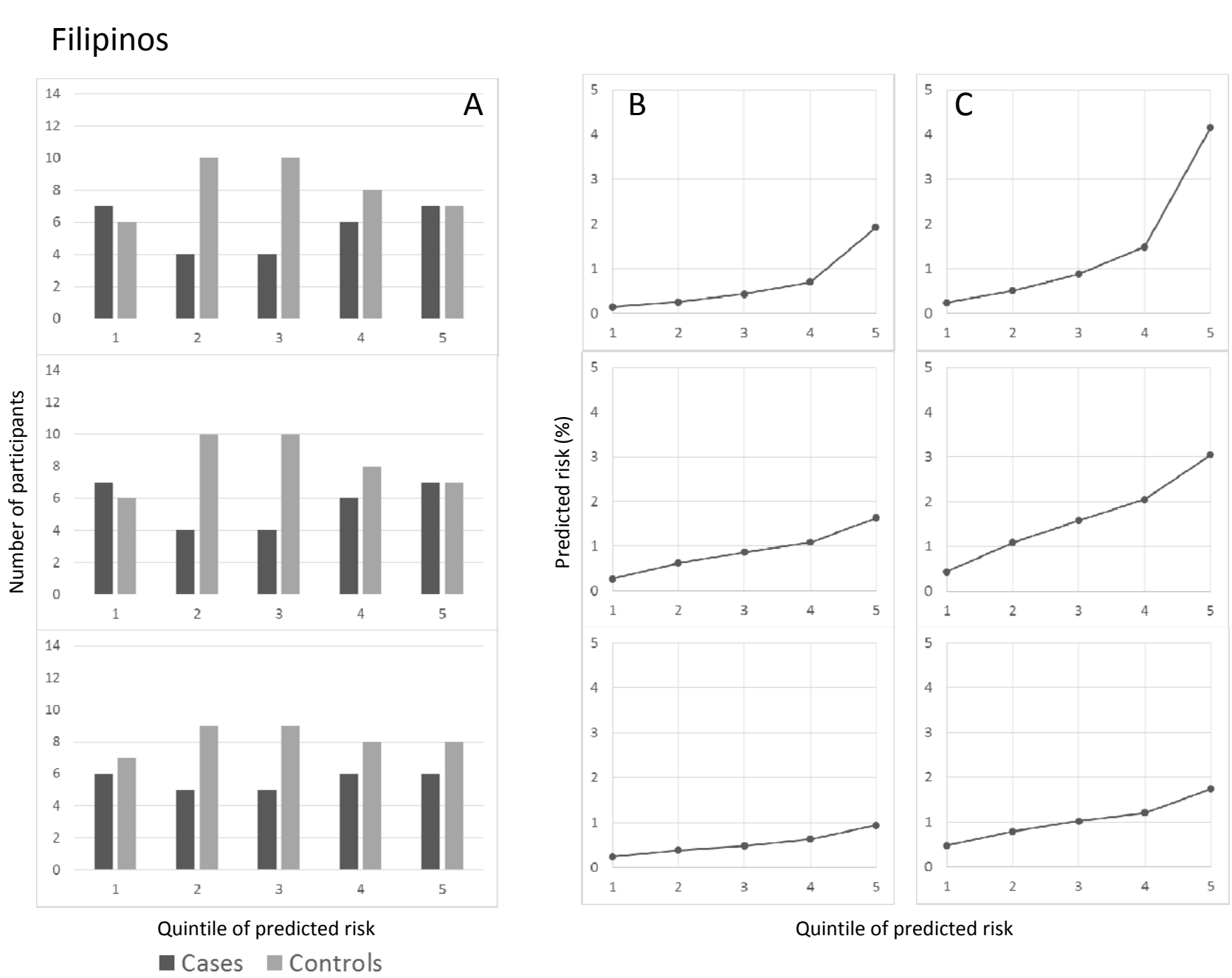
33. Bandera EV, Maskarinec G, Romieu I, *et al.* Racial and Ethnic Disparities in the Impact of Obesity on Breast Cancer Risk and Survival: A Global Perspective. Adv Nutr. 2015;6(6):803-19.
34. Nagrani R, Mhatre S, Rajaraman P, *et al.* Central obesity increases risk of breast cancer irrespective of menopausal and hormonal receptor status in women of South Asian Ethnicity. Eur J Cancer. 2016;66:153-61.
35. White KK, Park SY, Kolonel LN, *et al.* Body size and breast cancer risk: the Multiethnic Cohort. Int J Cancer. 2012;131(5):E705-16.
36. Chen J, Pee D, Ayyagari R, *et al.* Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. J Natl Cancer Inst. 2006;98(17):1215-26.
37. Maas P, Barrdahl M, Joshi AD, *et al.* Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA Oncol. 2016;2(10):1295-1302.

Figure 1. Cumulative incidence rates of invasive breast cancer in Guam and the U.S., 2000-2009.



Sources: (1) Guam Cancer Registry; (2) Hawaii Tumor Registry; (3) Surveillance, Epidemiology and End Results (SEER) 18-registry data.





Supplementary Table S1. Invasive breast cancer incidence rates in Guam women, 2000-2009 averages.

Age groups	Rate per 100,000	
	Filipino	Chamorro
20-24	0.0000	0.0000
25-29	6.8880	0.0000
30-34	19.8226	25.5028
35-39	87.2228	24.0964
40-44	76.7575	90.3465
45-49	137.0822	120.2289
50-54	110.8134	202.0321
55-59	186.8822	309.1133
60-64	85.5405	235.3095
65-69	139.3241	337.4036
70-74	227.2385	271.7201
75-79	123.1227	464.6494
80-84	108.3348	390.1445
>85	307.4785	564.2822

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *case-control studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(a) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	6
		(*b*) For matched studies, give matching criteria and the number of controls per case	6-7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-8
Bias	9	Describe any efforts to address potential sources of bias	6
Study size	10	Explain how the study size was arrived at	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-10
		(*b*) Describe any methods used to examine subgroups and interactions	8-10
		(*c*) Explain how missing data were addressed	8
		(*d*) If applicable, explain how matching of cases and controls was addressed	8
		(*e*) Describe any sensitivity analyses	9-10
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	7
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10-12
		(b) Indicate number of participants with missing data for each variable of interest	12

Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure	10-12
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	12-13
		(*b*) Report category boundaries when continuous variables were categorized	10-11
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	12
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12-13
Discussion			
Key results	18	Summarise key results with reference to study objectives	14
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	16
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	16
Generalisability	21	Discuss the generalisability (external validity) of the study results	14, 16
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17-18

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

TRIPOD Checklist: Prediction Model Development

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2-3
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	4-5
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	5
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	6-7
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	6
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	6
	5b	Describe eligibility criteria for participants.	6
	5c	Give details of treatments received, if relevant.	N/A
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	5, 6
	6b	Report any actions to blind assessment of the outcome to be predicted.	N/A
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	6-7
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	Explain how the study size was arrived at.	7
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	8
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	8-9
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	8-9
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	9-10
Risk groups	11	Provide details on how risk groups were created, if done.	9
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	10
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	10-11
Model development	14a	Specify the number of participants and outcome events in each analysis.	10
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	12-13
	15b	Explain how to the use the prediction model.	9, 11
Model performance	16	Report performance measures (with CIs) for the prediction model.	12-13
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	16-17
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	17
Implications	20	Discuss the potential clinical use of the model and implications for future research.	15-16
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	18
Funding	22	Give the source of funding and the role of the funders for the present study.	18

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

BMJ Open

Prediction of breast cancer risk among women of the Mariana Islands: the BRISK retrospective case-control study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-061205.R2
Article Type:	Original research
Date Submitted by the Author:	02-Nov-2022
Complete List of Authors:	Shvetsov, Yurii; University of Hawai'i at Mānoa, Cancer Center Wilkens, Lynne ; University of Hawai'i at Mānoa, Cancer Center White, Kami; University of Hawai'i at Mānoa, Cancer Center Chong, Marie; University of Hawai'i at Mānoa, Cancer Center Buyum, Arielle; AB Consulting, LLC Badowski, Grazyna; University of Guam, College of Natural and Applied Sciences Leon Guerrero, Rachael; University of Guam, College of Natural and Applied Sciences Novotny, Rachel; University of Hawai'i at Manoa, College of Tropical Agriculture and Human Resources
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Oncology, Research methods
Keywords:	Breast tumours < ONCOLOGY, Epidemiology < ONCOLOGY, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Prediction of breast cancer risk among women of the Mariana Islands: the BRISK retrospective case-control study

Yurii B. Shvetsov[1], Lynne R. Wilkens[1], Kami K. White[1], Marie Chong[1], Arielle Buyum[2], Grazyna Badowski[3], Rachael T. Leon Guerrero[4], Rachel Novotny[5]

[1]University of Hawaii Cancer Center, Honolulu, HI

[2]AB Consulting, LLC, Saipan, Northern Mariana Islands

[3]College of Natural & Applied Sciences, University of Guam, Mangilao, Guam

[4]Office of Research & Sponsored Programs, University of Guam, Mangilao, Guam

[5]College of Tropical Agriculture and Human Resources, University of Hawaii at Manoa, Honolulu, HI

Corresponding Author: Yurii B. Shvetsov, PhD, University of Hawaii Cancer Center, 701 Ilalo St, Honolulu HI 96813; tel: 808-564-5825; fax: 808-586-2982; email: yshvetso@cc.hawaii.edu.

Running Title: Prediction of breast cancer risk for Mariana Island women

Keywords: cancer: breast; cancer epidemiology; modeling

Financial Support: This work was supported by the U.S. National Cancer Institute, Comprehensive Partnerships to Reduce Cancer Health Disparities grants U54-CA143727 and U54-CA-143738, and by the U.S. National Cancer Institute grant R21-CA-220080.

Conflict of Interest: The authors declare no potential conflicts of interest.

Word count: 3355

Tables: 2

Figures: 3

ABSTRACT

Objectives: To develop a breast cancer risk prediction model for Chamorro and Filipino women of the Mariana Islands and compare its performance to that of the Breast Cancer Risk Assessment Tool (BCRAT).

Design: Case control study.

Setting: Clinics/facilities and other community-based settings on Guam and Saipan (Northern Mariana Islands).

Participants: 245 women (87 breast cancer cases and 148 controls) of Chamorro or Filipino ethnicity, age 25-80 years, with no prior history of cancer (other than skin cancer), residing on Guam or Saipan for at least 5 years.

Primary and secondary outcome measures: breast cancer risk models were constructed using combinations of exposures previously identified to affect breast cancer risk in this population, population breast cancer incidence rates and all-cause mortality rates for Guam.

Results: Models utilizing ethnic-specific relative risks performed better than those with relative risks estimated from all women. The model with the best performance among both ethnicities (the BRISK model; AUC: 0.66 and 0.65 among Chamorros and Filipinos, respectively) included age at first live birth and waist circumference. The 10-year breast cancer risk predicted by the BRISK model was 1.36% for Chamorros and 0.93% for Filipinos. Performance of the BCRAT was modest among both Chamorros (AUC: 0.60) and Filipinos (AUC: 0.55), possibly due to incomplete information on BCRAT risk factors.

Conclusions: The ability to develop breast cancer risk models for Mariana Islands women is constrained by the small population size and limited availability of health services and data. Nonetheless, we have demonstrated that breast cancer risk prediction models with adequate discriminatory performance can be built for small populations such as in the Mariana Islands. Anthropometry, in particular waist circumference, was important for estimating breast cancer risk in this population.

ARTICLE SUMMARY

Strengths and limitations of the study

- The small sample size of this study is a direct consequence of the small population size.
- Our model construction method is designed to overcome the challenge of small population size.
- Bootstrap validation was used to minimize optimism bias.
- Evaluation of model coefficients separately for Chamorro and Filipino women of the Mariana Islands accounted for possible differential effect of model predictors between these two ethnic groups.

INTRODUCTION

Breast cancer is the most common cancer among women worldwide.[1] It is the second most common cause of cancer mortality among U.S. women [2] and has been the leading cause of cancer mortality among women on Guam over the last three decades.[3]

The Mariana Islands consist of two administrative units: Guam, a U.S. territory, and the Commonwealth of the Northern Mariana Islands (CNMI), which includes the islands of Saipan, Tinian, and Rota. The current population of Guam is ethnically mixed,[4] with 37% Chamorro, 26% Filipino, 12% other Pacific Islander, and 25% other ethnicity. CNMI is also diverse; its ethnic breakdown includes 24% Chamorro, 35% Filipino, 11% other Pacific Islander, and 30% other ethnicity.[5]

While the breast cancer incidence rate on Guam is lower than across the U.S., breast cancer mortality among some ethnicities on Guam, especially Chamorros, is higher than among U.S. women.[6] During 1998–2002 on Guam, the age-adjusted breast cancer incidence rate among Chamorro women was nearly twice as high as Filipino women and second only to White women (115.9, 60.7 and 148.6 per 100,000, respectively).[7] The age-adjusted incidence rate for U.S. women (not including data from the US affiliated Mariana Islands) during this time was 131 per 100,000

women. Chamorro women also had the highest breast cancer mortality rate on Guam, at 32 per 100,000 women.[8] This contrasts with the overall U.S. mortality rate for that time period of 28 per 100,000.

The reasons for higher breast cancer mortality rates, and relatively high incidence rates, among Chamorro Pacific Islanders compared with other ethnic groups in the Mariana Islands are not well understood. The Breast Cancer Risk Model (BRISK) Project was conducted to improve understanding of the risk factors for breast cancer in this region.[9]

Estimation of a woman's breast cancer risk is an important tool used for risk assessment and stratification in breast cancer screening and prevention efforts. One of the most widely used models for predicting breast cancer risk is the Gail model, developed for white women [10, 11] and subsequently extended to include other race/ethnicities such as African American and Asian American women.[12, 13] This extended model is available as NCI's Breast Cancer Risk Assessment Tool (BCRAT).[14] Although BCRAT includes Filipinos as one of the Asian American ethnicities, it is built from the Filipino population in SEER 9 registries,[15] whose age-specific breast cancer incidence rates differ from those for Filipinos on Guam, a US territory (Figure 1). A similar situation exists for Pacific Islanders, where only rates for Native Hawaiians are present in BCRAT. Additionally, BCRAT uses the same risk factors and relative risk estimates for all Asian American ethnicities; however, different breast cancer risk models are needed for adequate risk estimation for women of diverse racial/ethnic backgrounds,[16] and while some of the established risk factors are associated with breast cancer risk in the Mariana Islands women, others are not.[9] Due to these considerations, the utility of the BCRAT model for the Mariana Islands women is unknown.

In the present report, we evaluate performance of the BCRAT model and its modified version among Chamorro and Filipino participants in the BRISK study. In so doing, we propose a method of

risk model development for small populations which we use here for the development and internal validation of a new breast cancer risk model for Chamorro and Filipino women of the Mariana Islands.

METHODS

BRISK study design and population

BRISK is a retrospective case-control study of mostly Asian and Pacific Islander women living on the Mariana Islands of Guam and Saipan. The study was a collaboration between the University of Guam and the University of Hawaii Cancer Center and was approved by the Institutional Review Boards at both institutions.

A detailed description of the study design and recruitment is provided elsewhere.[9, 17] Briefly, breast cancer cases and controls were recruited between 2010 and 2013. Breast cancer cases were identified through the Guam Cancer Registry, CNMI Department of Public Health, and health clinics on Guam. Controls were recruited in local clinics/facilities and other community-based settings on Guam and Saipan from among women with mammography screening and were frequency-matched to cases on age, ethnicity, and location (Saipan or Guam). Eligibility criteria for all participants were: (1) no prior history of cancer (other than skin cancer); (2) residence on Guam or Saipan for at least 5 years; (3) ability to provide consent for the study; and (4) age between 25 and 80 years. An additional eligibility criterion for cases was primary, invasive breast cancer newly diagnosed between 2009 and 2012.

During an interview, participants completed a detailed questionnaire including demographic, anthropometric, behavioral and lifestyle information; personal and family medical history; reproductive history; and acculturation based on a survey used in a multiethnic study.[18, 19] The reference date for the interview was the diagnosis date for cases and the interview date for controls. In addition, current waist circumference (WC), measured with an inelastic tape measure at the level of the

umbilicus,[20] weight, height, and sitting height were measured by a trained anthropometrist. Body mass index (BMI) was calculated as kg/m^2. Waist-height ratio (WHtR) was calculated as WC in cm divided by height in cm.

Of the 275 cases contacted, 38% agreed to participate, 21% were ineligible, and 41% refused due to scheduling conflicts, lack of transportation, family, psychological or cultural reasons, or off-island travel.[21] The corresponding percentages for controls were 74%, 20% and 6%. The study included 104 breast cancer cases (83 from Guam and 21 from CNMI) and 185 controls (140 from Guam and 45 from CNMI) between 27 and 80 years of age. A summary ethnicity variable was defined based on each participants' self-reported composition of her mother's and father's ethnicities. The present analysis was limited to participants with summary ethnicity of Chamorro and Filipino residing on Guam and Saipan (87 cases and 158 controls).

Patient and Public Involvement

Patients were not involved in the development of the research question, design of the study, recruitment and conduct of the study. However, the study provided funds to the CNMI Public Health mammography program to expand access and facilitate recruitment. The results were disseminated to study participants by public talks given at the University of Guam.

Breast cancer incidence and all-cause mortality rates

We obtained data from the Guam Cancer Registry (GCR) for all reportable female breast cancer diagnoses (n=576) on Guam for 2000-2009 (Supplementary Table S1).[17] Since data for CNMI were unavailable, Guam rates were also used to represent Saipan. Average annual age-specific incidence rates for female breast cancer were computed per ethnicity and 5-year age group, using interpolations between the U.S. 2000 and 2010 female census counts for Guam as denominators. All-cause mortality

rates were obtained from the Guam Statistical Yearbook 2004.[22] Since 2004 was the only year these rates were published, the rates for 2004 were used as a reasonable approximation for the 2000-2009 all-cause mortality rates.

Construction and selection of risk models

We assumed the general form of the Gail model,[10, 13, 23, 24] which projects absolute risk of breast cancer at a specified time interval using relative risk estimates for a set of risk factors, population breast cancer incidence rates and all-cause mortality rates. Risk factors considered for inclusion in the models were those identified in our previous report [9] as having a statistically significant ($p < 0.05$) association with breast cancer risk among Guam and Saipan women: age at first live birth (<20 or missing, 20-24, 25-29 or nulliparous, ≥30 y); BMI (<25, 25-29, ≥30); WHtR (≤0.54, 0.55-0.61 or missing, 0.62-0.67, >0.67); and WC (≤89, 90-99.5 or missing, >99.5 cm). Also considered for inclusion were the risk factors included in the original Gail model [10, 13] although they did not have a statistically significant association with breast cancer risk in our study: age at menarche (<12, 12-13, ≥14 y or missing); first-degree relatives with breast cancer (yes, no) and menopausal status (premenopausal, postmenopausal). As BMI, WHtR and WC were strongly correlated in our study, only one of these 3 factors was allowed to enter the model at a time. Following the approach of Gail et al.,[10] for each risk factor, missing values were grouped with the category showing the closest risk of breast cancer to participants with missing values, according to minimally adjusted logistic models. We constructed and evaluated models that included every combination of the above 7 risk factors as main effects (a total of 127 models). For each such combination, the entire dataset was used to estimate odds ratios (ORs) for the included risk factors using multivariable unconditional logistic regression, with adjustment for study participants' age, among both ethnicities combined and separately for Chamorros and Filipinos. Model-based adjusted attributable risk (AR) corresponding to these risk factors was then

computed.[25] The Hosmer-Lemeshow statistic was computed to assess model fit. A risk model was constructed using the OR and AR estimates from the logistic model. To assess model performance, a bootstrap validation method was utilized, whereby a validation subset was randomly selected, containing 50% of breast cancer cases (n=42) and two age and ethnicity-matched controls per case. The model was applied to all participants in the validation subset to project the absolute risk of breast cancer for a five-year period preceding the study interview date, and the area under the ROC curve (AUC) statistic was computed.

This bootstrap validation step was performed 100 times for each model, and the median AUC was computed. The top performing BRISK model was selected based on the highest median AUC for each ethnicity.

Evaluation of model performance

The final BRISK model was examined for its calibration and discrimination. The median AUC across bootstrap validation steps and its 95% confidence interval were taken as the measure of discriminatory performance of the model. Calibration of the model was assessed by examining the case/control distribution within quintiles of predicted 5- and 10-year absolute risk across the entire sample. The mean predicted risk of breast cancer was also computed for each quintile. Performance was compared with that of BCRAT.[13] As Native Hawaiians are the only Pacific Islander ethnicity represented in BCRAT and are the closest to Chamorros in terms of culture and lifestyle, we used Native Hawaiian incidence and mortality rates when applying BCRAT to Chamorro women. Due to a lack of breast biopsy information in our sample, all women were assumed to have had no breast biopsies, the default value in BCRAT.

Additionally, to examine whether calibrating the BCRAT model to the Guam breast cancer incidence rates would improve its performance, we modified the BCRAT model by replacing incidence

and mortality rates with those for Filipino and Chamorro women on Guam, while retaining risk factors and their relative risk estimates specified in the BCRAT; this modified model is referred to as BCRAT-G.

RESULTS

The demographic, lifestyle and reproductive characteristics of the study participants included in the present analysis (n=245) are summarized in Table 1. Briefly, the largest age group among both cases and controls was 50-59 years. One third of the participants (33%) were of Filipino ethnicity, the rest were Chamorros. The ethnic composition was similar among cases and controls by design, although the case to control ratio was somewhat higher among Filipino than Chamorro women (43% and 32% cases, respectively). Cases and controls had a similar proportion of women ever pregnant, pre-menopausal, parous, and having ever breastfed, but somewhat differed in BMI, WC, WHtR, alcohol consumption and smoking.

Table 1. Characteristics[a] of breast-cancer cases and controls among Chamorro and Filipino women of Mariana Islands in the BRISK study.

Characteristic, n (%)	Cases (n = 87)	Controls (n = 158)	P-value[b]
Age at reference[c], years (mean ± SD)	55.1 ± 10.8	53.8 ± 10.6	0.35
<40	7 (8.1)	12 (7.6)	0.92
40–49	22 (25.3)	47 (29.7)	
50–59	29 (33.3)	54 (34.2)	
60–69	19 (21.8)	31 (19.6)	
≥70	10 (11.5)	14 (8.9)	
Ethnicity			0.11
Chamorro	53 (60.9)	112 (70.9)	
Filipino	34 (39.1)	46 (29.1)	
Highest education level completed			0.99
High school diploma or less	40 (46.0)	73 (46.2)	
Some college	25 (28.7)	46 (29.1)	
College degree or more	22 (25.3)	39 (24.7)	
Age at menarche, years[c]			0.39
<12	20 (23.0)	45 (28.9)	

12–13	35 (40.2)	66 (42.3)	
≥14	32 (36.8)	45 (28.9)	
Ever been pregnant	78 (89.7)	145 (91.8)	0.58
Total number of pregnancies			0.67
0	9 (10.3)	13 (8.2)	
1–2	26 (29.9)	39 (24.7)	
3–4	31 (35.6)	59 (37.3)	
5 or more	21 (24.1)	47 (29.8)	
Number of live births			0.17
Nulliparous	10 (11.5)	17 (10.8)	
1–2	36 (41.4)	45 (28.5)	
3–4	25 (28.7)	63 (39.9)	
5 or more	16 (18.4)	33 (20.9)	
Age at first live birth, years, parous women only (mean ± SD)[c]	25.0 ± 5.5	22.9 ± 5.2	0.006
<20	18 (23.4)	48 (34.5)	0.03
20–24	22 (28.6)	52 (37.4)	
25–29	25 (32.5)	26 (18.7)	
≥30	12 (15.6)	13 (9.4)	
Ever breastfed, parous women only			0.83
No	24 (31.2)	42 (29.8)	
Yes	53 (68.8)	99 (70.2)	
Number of first-degree relatives with breast cancer			0.39
0	76 (87.4)	132 (83.5)	
1	8 (9.2)	23 (14.6)	
2	3 (3.4)	3 (1.9)	
Hormone use[c]			0.35
Never used estrogen or progesterone	77 (90.6)	133 (84.7)	
Yes, previously	8 (9.4)	21 (13.4)	
Yes, currently	0	3 (1.9)	
Menopausal status			0.21
Premenopausal	25 (28.7)	48 (30.4)	
Perimenopausal	4 (4.6)	17 (10.8)	
Postmenopausal	58 (66.7)	93 (58.9)	
Body mass Index, kg/m^2 (mean ± SD)	29.8 ± 7.0	30.3 ± 7.4	0.55
<18	0	0	0.10
18–24.9	18 (20.7)	44 (27.9)	
25–29.9	35 (40.2)	49 (31.0)	
≥30	34 (39.1)	65 (41.1)	
Waist Circumference, cm (mean ± SD)	97.2 ± 14.6	94.5 ± 14.9	0.19
Tertile 1 (≤89)[d]	24 (30.0)	53 (36.3)	0.54
Tertile 2 (89.1–99.5)	28 (35.0)	51 (34.9)	
Tertile 3 (>99.5)	28 (35.0)	42 (28.8)	
Waist/Height Ratio (mean ± SD)[e]	0.63 ± 0.09	0.61 ± 0.10	0.31
Quartile 1 (≤0.54)[d]	13 (16.3)	35 (24.0)	0.58
Quartile 2 (0.55–0.62)	27 (33.8)	43 (29.5)	
Quartile 3 (0.62–0.67)	20 (25.0)	32 (21.9)	
Quartile 4 (>0.67)	20 (25.0)	36 (24.7)	
Alcohol intake, drinks/week[e]			0.04
None	48 (76.2)	87 (61.7)	
Any alcohol reported	15 (23.8)	54 (38.3)	

Smoked daily for > 6 months[e]			0.05
No	61 (70.9)	92 (58.2)	
Yes	25 (29.1)	66 (41.8)	

[a]Percentage is based on non-missing data and may not add up to 100 due to rounding.

[b]P-values based on chi-square test for categorical characteristics and t-test for continuous characteristics.

[c]Reference date was defined as diagnosis date for cases, interview date for controls.

[d]Quartiles and tertiles are based on the distribution among both cases and controls.

[e]Missing values were excluded: 2 controls for age at menarche, 2 controls for age at first live births, 2 cases and 1 control for hormone use, 7 cases and 12 controls for waist/height ratio, 24 cases and 17 controls for alcohol intake, 1 case for smoked daily for >6 months.

The composition of the top BRISK model and its performance are summarized in Table 2. The model included separate relative risk estimates among Chamorros and Filipinos for the included risk factors: age at menarche, age at first live birth and the number of first-degree relatives with breast cancer for both ethnicities, and additionally WC for Filipino women. The AUCs among Chamorros and Filipinos, respectively, were 0.64 and 0.67, based on the median across 100 validation runs.

The BRISK model classified more cases than controls into the highest risk stratum and more controls than cases into the lowest risk stratum among both ethnicities (Figures 2, 3), which indicates a good performance in terms of case/control distribution. Using case and control data, the BRISK model predicted a median 10-year absolute risk of breast cancer to be 1.28% for Chamorro women and 0.89% for Filipino women.

Table 2. Performance of the BRISK model and BCRAT among Mariana Island women in the BRISK study.

	BRISK[1,2]		BCRAT[3]		BCRAT-G[4]	
	Chamorros	Filipinos	Chamorros	Filipinos	Chamorros	Filipinos
Risk factors included / odds ratios:						
Age at menarche	1.13 (0.72-1.78)	1.71 (0.84-3.49)	1.078		1.078	
Age at first live birth	1.79 (1.25-2.56)	0.91 (0.58-1.41)	1.318		1.318	

Waist circumference	---	1.97 (1.19-3.25)	---		---	
Number of relatives with breast cancer	0.96 (0.39-2.36)	0.61 (0.09-4.07)	2.207		2.207	
Number of biopsies[5]	---	---	1.738		1.738	
Hosmer-Lemeshow statistic p-value[6]	0.52	0.86				
AUC (95% CI)[7]	0.64 (0.63 - 0.65)	0.67 (0.65 - 0.68)	0.60 (0.50 - 0.69)	0.55 (0.40 - 0.70)	0.59 (0.49 - 0.69)	0.51 (0.36 - 0.66)
Difference (% risk) in the median estimated risk between cases and controls[7]	0.33 (0.27-0.38)	0.31 (0.28-0.36)	0.18	0.13	0.06	0.00

[1]Highest AUC among Chamorros and Filipinos.

[2]Odds ratios for included risk factors are estimated in BRISK separately for Chamorros and Filipinos.

[3]BCRAT absolute risk estimates, selecting the Native Hawaiian BCRAT model for Chamorros.

[4]BCRAT absolute risk estimates, selecting the Native Hawaiian BCRAT model for Chamorros, with substitution of the breast cancer incidence and mortality rates by Guam rates.

[5]Number of biopsies was not available in our study and therefore was assigned the default value in the models.

[6]Computed using the underlying logistic regression model.

[7]Estimated as the median from 100 bootstrap validation datasets (30% data) for the BRISK model. Estimated using all data for BCRAT and BCRAT-G.

OR: odds ratio. AUC: area under the receiver operating characteristic curve. CI: confidence interval.

The unmodified BCRAT and the modified BCRAT-G model exhibited similar performance among Chamorros (AUC: 0.60 and 0.59, respectively) while BCRAT performed non-significantly better than BCRAT-G among Filipinos (AUC: 0.55 and 0.51, respectively; Table 2). Both models performed better among Chamorros than among Filipinos. Both BCRAT and BCRAT-G classified more controls than cases into the lower risk stratum among Filipinos, but not among Chamorros (Figures 2, 3). Both BCRAT and BCRAT-G classified more cases than controls into the higher risk stratum among both Chamorros and Filipinos.

DISCUSSION

To our knowledge, this is the first study that tested existing, as well as developed new, breast cancer risk models in a small, isolated population such as the Mariana Islands and in Pacific Islander populations other than Native Hawaiians. Developing or validating cancer risk models for populations such as Mariana Islands is challenging. Due to its unique ethnic composition and lifestyle, this population may be subject to unique risk factors not affecting other populations. The small population size places a natural restriction on the sample size of any epidemiologic study and reduces statistical power for potential model development. The population's geographic isolation results in the absence of sufficiently large comparable populations for external model validation.

A key challenge in our study was its small sample size, largely precipitated by the small size of the target population and newly emerging breast cancer registries. It is generally recommended that any new risk prediction model should include internal validation, either as bootstrap validation or utilizing training and validation subsets.[26, 27] As splitting a small dataset into training and validation parts would cause instability in the relative risk estimates and consequently in the resulting model, we have implemented a bootstrap validation procedure and used the entire dataset for parameter estimation. Our method produced a model that performed reasonably well, with AUC of 0.64-0.67 comparable to the AUC range of 0.53-0.68 for other published models.[28, 29] We also found that performance of the BCRAT model was modest among Chamorro and Filipino women in our study, with AUCs not exceeding 0.60. The poor performance of BCRAT-G indicates that replacing population incidence and mortality curves with those from the target population did not improve model performance.

There are several possible reasons that could explain the observed differences in model performance. First, in addition to the established risk factors in the Gail model, only the risk factors that exhibited significant associations with breast cancer risk in BRISK were considered for inclusion in the development of the model. Including risk factors not significantly associated with the outcome may cause model overfitting,[30] which may in turn bias the predicted absolute risk. In our previous

report,[9] no significant association between breast cancer risk and a number of known risk factors was found, but significant effects of several anthropometric factors such as BMI, WC and WHtR were observed on the risk of breast cancer. This may indicate a unique risk profile for this population or minimal variation in the known risk factors.

The BRISK model utilizes separate relative risk estimates by ethnicity with an additional risk factor (WC) for Filipinos; the model utilizing joint estimates did not perform as well. This indicates that Chamorros and Filipinos have different breast cancer risk profiles, which should be taken into account in risk prediction models. The BRISK model included anthropometrics in the form of WC, which reinforces the need to consider anthropometric measures in breast cancer risk models. Body size is dramatically different among the Asian and Pacific Islander residents in the Mariana Islands, with Filipino women generally having smaller body size than Chamorro women.[31, 32] BMI and central obesity have been found to be associated with higher breast cancer risk among Asian women,[33-35] and studies have demonstrated that the addition of body size variables improves prediction of breast cancer risk.[36] The inclusion of WC for Filipinos only may have to do with the issue of differing body sizes and excess overweight/obesity rates among Chamorros, thus diminishing the predictive value of body size for breast cancer in this ethnic group.

The BRISK model included 3-4 risk factors out of seven considered for inclusion. It has been suggested that the complexity threshold for a risk prediction model is 20 cases per model parameter.[26, 30] Exceeding this threshold in terms of the number of model parameters increases the danger of overfitting. In our study, with 87 breast cancer cases, the optimal number of model parameters is <5, which is evidenced in the final model. Applying a similar method of model selection and validation to a larger dataset may have resulted in a model with more parameters.

A recent focus of the breast cancer risk model improvement efforts has been examination of modifiable risk factors and their impact on predicted breast cancer risk.[37] The BRISK model

includes WC, a modifiable factor. This opens the possibility of the model being used as a supplemental health assessment tool in health behavior interventions, providing additional motivation for adoption of a healthier lifestyle that could decrease WC. As all predictors in the model can be collected from a patient questionnaire, the model can easily be implemented in most clinic settings including local clinics.

Limitations of our study include the small sample size as noted above, which may have prevented us from detecting important risk factors and, combined with limited response rate, may limit generalizability of findings. The failure to detect some expected associations (and thus to include the corresponding risk factors in the models) may also be due to a small sample size and lack of variability of some exposures in the study sample. The information on risk exposures was limited; in particular, performance of the BCRAT model could have been affected by the lack of information on breast biopsies in our study. The Guam breast cancer incidence rates covered a 10-year period and, thus, can be deemed reliable; however the all-cause mortality rates in our study are based on one year and thus may not be sufficiently stable. No CNMI breast cancer incidence or mortality rates were available and had to be approximated by the Guam rates. Although the ethnic composition of CNMI is similar to Guam, it is possible that, due to differing lifestyle and poorer access to healthcare among the CNMI population, the breast cancer incidence and overall mortality rates in CNMI differ from those on Guam. However, since the majority of study participants were from Guam, this potential difference in rates was unlikely to have a substantial impact on our results. Nonetheless, efforts are needed to collect and disseminate data on cancer incidence and mortality rates for CNMI, which would allow researchers to improve study results.

Because BRISK was a case-control study, we were unable to assess model calibration to population incidence rates, although we examined the internal calibration of the model. We note, however, that the AUC-based comparison of models is robust to mis-calibration [24] and thus is a valid

method in our study. We were also unable to perform external validation of the BRISK model, which is challenging given the unique nature and small size of this population, and remains a topic for future studies. Finally, AUCs based on the same dataset used for model construction may be overly optimistic.[30] We used the bootstrap validation method to minimize the optimism bias, although some of it may still persist. Despite these limitations, our model construction method has produced a reasonably well performing breast cancer risk model for Chamorro and Filipino women of the Mariana Islands, and the first and only model for this population.

CONCLUSIONS

We have demonstrated that breast cancer risk prediction models with adequate discriminatory performance can be built for small populations such as the Mariana Islands. The proposed model has the potential of being useful as a supplemental tool for risk assessment and stratification in breast cancer screening and prevention in the Mariana Islands, but needs further refinement on larger samples of women and external validation on comparable Pacific Island populations.

ACKNOWLEDGEMENTS

We thank Michelle Blas-Laguana, Ashley Yamanaka, and Frances Santos-Hofschneider for recruiting and interviewing the BRISK Project participants, and the staff of Guam Radiology Consultants, FHP Clinic, and Guam Seventh-Day Adventist Clinic for their assistance in the recruitment of participants. We also thank the participants on Guam and Saipan who volunteered to take part in the BRISK study.

AUTHORS' CONTRIBUTIONS

Y.B.S. conducted the primary statistical analysis and had primary responsibility for the final manuscript; R.T.L.G. and R.N. led study concept and design; L.R.W. ensured integrity and accuracy of

the study data; A.B. led data collection in Saipan; Y.B.S., L.R.W., K.K.W., M.C., G.B. contributed to statistical analysis; Y.B.S., L.R.W., K.K.W., R.T.L.G., R.N. interpreted the results and wrote the manuscript; Y.B.S., L.R.W., K.K.W., A.B., R.T.L.G., R.N. reviewed and approved the final manuscript.

FUNDING

This work was supported by the U.S. National Cancer Institute, Comprehensive Partnerships to Reduce Cancer Health Disparities grants U54-CA143727 and U54-CA-143738, and by the U.S. National Cancer Institute grant R21-CA-220080.

DATA SHARING

The datasets generated and analyzed during the current study are not publicly available because they contain protected health information. De-identified datasets are available from the senior author (R.N. at novotny@hawaii.edu) on reasonable request.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study has received ethical approval from the Institutional Review Boards at the University of Guam (ref. number 0982) and the University of Hawaii (ref. number 17796). Written informed consent to participate was obtained from all study participants. All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

COMPETING INTERESTS

The authors declare that they have no competing interests to disclose.

FIGURE LEGEND

Figure 1. Cumulative incidence rates of invasive breast cancer in Guam and the U.S., 2000-2009.

Figure 2. Classification of breast cancer cases and controls into risk strata among Chamorro women in the BRISK study: the BRISK Model, BCRAT and BCRAT-G. A: Frequency of cases and controls by quintile of predicted risk. B: Mean predicted 5-year risk by quintile of predicted risk. C: Mean predicted 10-year risk by quintile of predicted risk.

Figure 3. Classification of breast cancer cases and controls into risk strata among Filipino women in the BRISK study: the BRISK Model, BCRAT and BCRAT-G. A: Frequency of cases and controls by quintile of predicted risk. B: Mean predicted 5-year risk by quintile of predicted risk. C: Mean predicted 10-year risk by quintile of predicted risk.

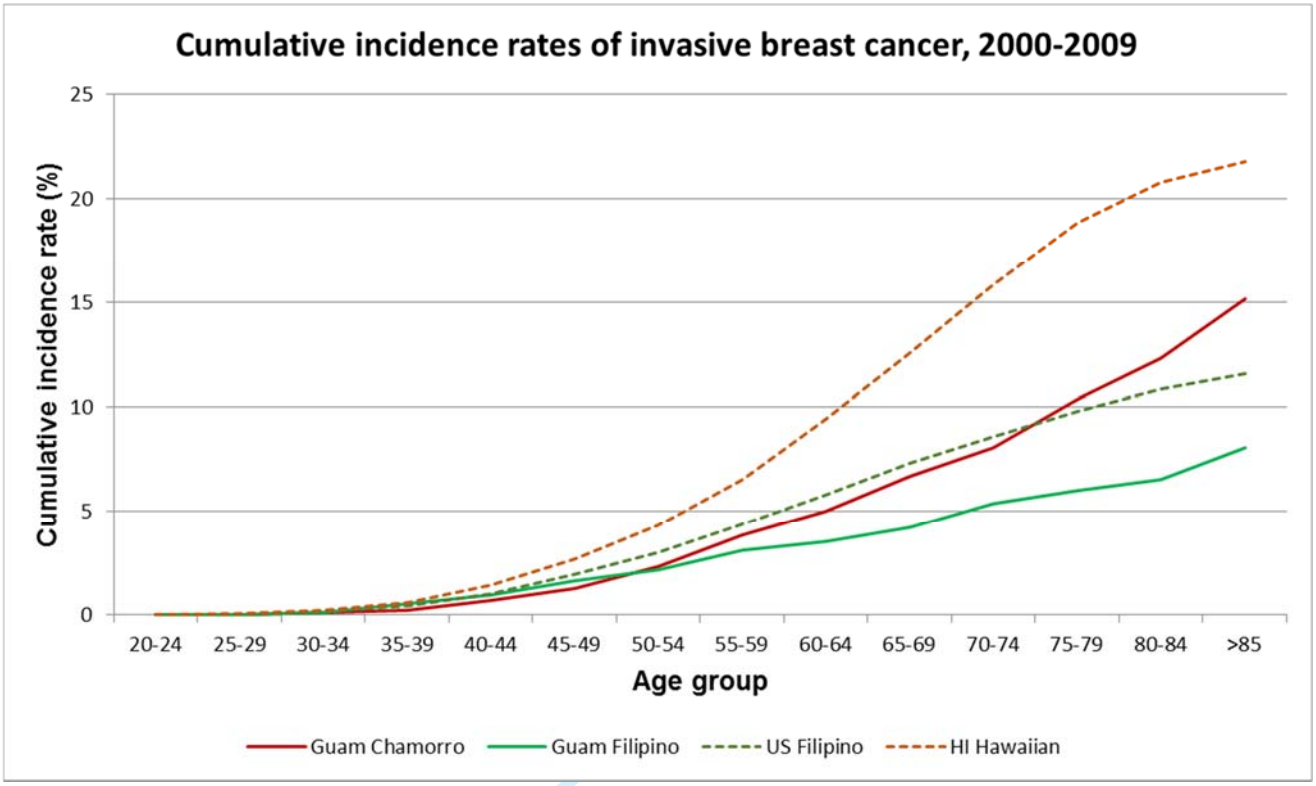
References

1. Ferlay J, Soerjomataram I, Dikshit R, *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359-86.
2. American Cancer Society. *Cancer facts & figures*. Atlanta, GA; 2018.
3. Haddock RL, Naval C. Cancer on Guam, especially among Micronesians. Pac Health Dialog. 2002;9:222–225.
4. Central Intelligence Agency. *The World Factbook*. https://www.cia.gov/library/publications/resources/the-world-factbook/geos/gq.html. Accessed November 19, 2018.
5. CNMI Department of Commerce. *Census demographics profile summary by district*. http://commerce.gov.mp/wp-content/uploads/2012/12/2010-Census-Demographics-Profile-Summary-by-District.pdf. Accessed November 19, 2018.
6. Haddock RL, Whippy HJ, Talon RJ, *et al.* Ethnic disparities in cancer incidence among residents of Guam. Asian Pac J Cancer Prev. 2009;10(1):57-62.
7. Guam Comprehensive Cancer Control Program. *Guam Cancer Facts & Figures 2008–2012*. Mangilao, GU; 2015.
8. Haddock RL, Talon RJ, Whippy HJ. Ethnic disparities in cancer mortality among residents of Guam. Asian Pac J Cancer Prev. 2006;7(3):411-4.
9. Leon Guerrero RT, Novotny R, Wilkens LR, *et al.* Risk factors for breast cancer in the breast cancer risk model study of Guam and Saipan. Cancer Epidemiol. 2017;50(Pt B):221-233.
10. Gail MH, Brinton LA, Byar DP, *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81(24):1879-86.

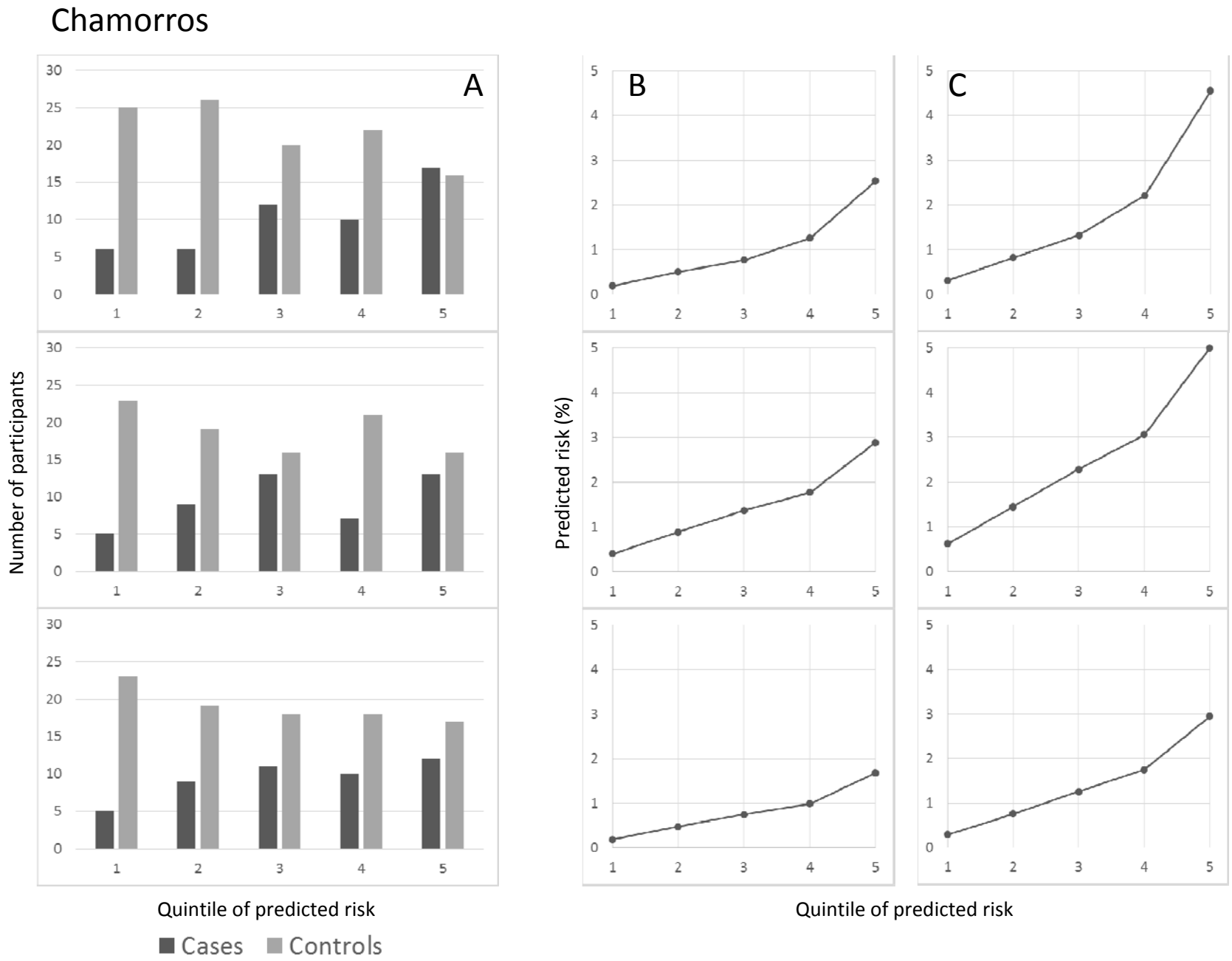
11. Costantino JP, Gail MH, Pee D, *et al.* Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst. 1999;91(18):1541-8.
12. Gail MH, Costantino JP, Pee D, *et al.* Projecting individualized absolute invasive breast cancer risk in African American women. J Natl Cancer Inst. 2007;99(23):1782-92.
13. Matsuno RK, Costantino JP, Ziegler RG, *et al.* Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. J Natl Cancer Inst. 2011;103(12):951-61.
14. *Breast Cancer Risk Assessment SAS Macro (Gail Model)*. http://dceg.cancer.gov/tools/risk-assessment/bcrasasmacro. Accessed September 5, 2017.
15. Surveillance E, and End Results (SEER) Program (www.seer.cancer.gov). SEER*Stat Database: Incidence - SEER Research Data, 9 Registries, Nov 2018 Sub (1975-2016) - Linked To County Attributes - Time Dependent (1990-2016) Income/Rurality, 1969-2017 Counties. In: National Cancer Institute D, Surveillance Research Program, (ed). November 2018 ed.
16. Bernstein L, Teal CR, Joslyn S, *et al.* Ethnicity-related variation in breast cancer risk factors. Cancer. 2003;97(1 Suppl):222-9.
17. Leon Guerrero RT, Badowski G, Yamanaka A, *et al.* University of Hawai'i Cancer Center connection: The vital role of cancer registries in the recruitment of an understudied minority population into a breast cancer study: Breast Cancer Risk Model for the Pacific. Hawaii J Med Public Health. 2014;73(10):335-40.
18. Goodman MT, Wu AH, Tung KH, *et al.* Association of dairy products, lactose, and calcium with the risk of ovarian cancer. Am J Epidemiol. 2002;156(2):148-57.
19. Tung KH, Goodman MT, Wu AH, *et al.* Reproductive factors and epithelial ovarian cancer risk by histologic type: a multiethnic case-control study. Am J Epidemiol. 2003;158(7):629-38.
20. Lohman TG, Roche AF, Martorell R. *Anthropometric standardization reference manual*. Champaign, IL: Human Kinetics Books; 1988.
21. Leon Guerrero RT, Badowski G, Yamanaka A, *et al.* University of Hawai'i Cancer Center connection: The vital role of cancer registries in the recruitment of an understudied minority population into a breast cancer study: Breast Cancer Risk Model for the Pacific. Hawaii J Med Public Health. 2014;73(10):335-340.
22. Guam Bureau of Statistics and Plans. *Guam Statistical Yearbook*. Mangilao, GU; 2004.
23. Gail MH, Costantino JP. Validating and improving models for projecting the absolute risk of breast cancer. J Natl Cancer Inst. 2001;93(5):334-5.
24. Pfeiffer RM, Gail MH. *Absolute Risk: Methods and Applications in Clinical Management and Public Health*. Boca Raton, FL: CRC Press; 2017.
25. Bruzzi P, Green SB, Byar DP, *et al.* Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol. 1985;122(5):904-14.
26. Dupont WD, Blume JD, Smith JR. Building and Validating Complex Models of Breast Cancer Risk. JAMA Oncol. 2016;2(10):1271-1272.
27. Collins GS, Reitsma JB, Altman DG, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Ann Intern Med. 2015;162(10):735-6.
28. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, *et al.* Risk prediction models of breast cancer: a systematic review of model performances. Breast Cancer Res Treat. 2012;133(1):1-10.
29. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat. 2012;132(2):365-77.
30. Harrell FE. *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
31. LeonGuerrero RT, Murphy SP, Novotny R, *et al.* Diet and obesity among Chamorro and Filipino Adults on Guam. Asia Pacific Journal of Clinical Nutrition. 2008;17(2):216-222.

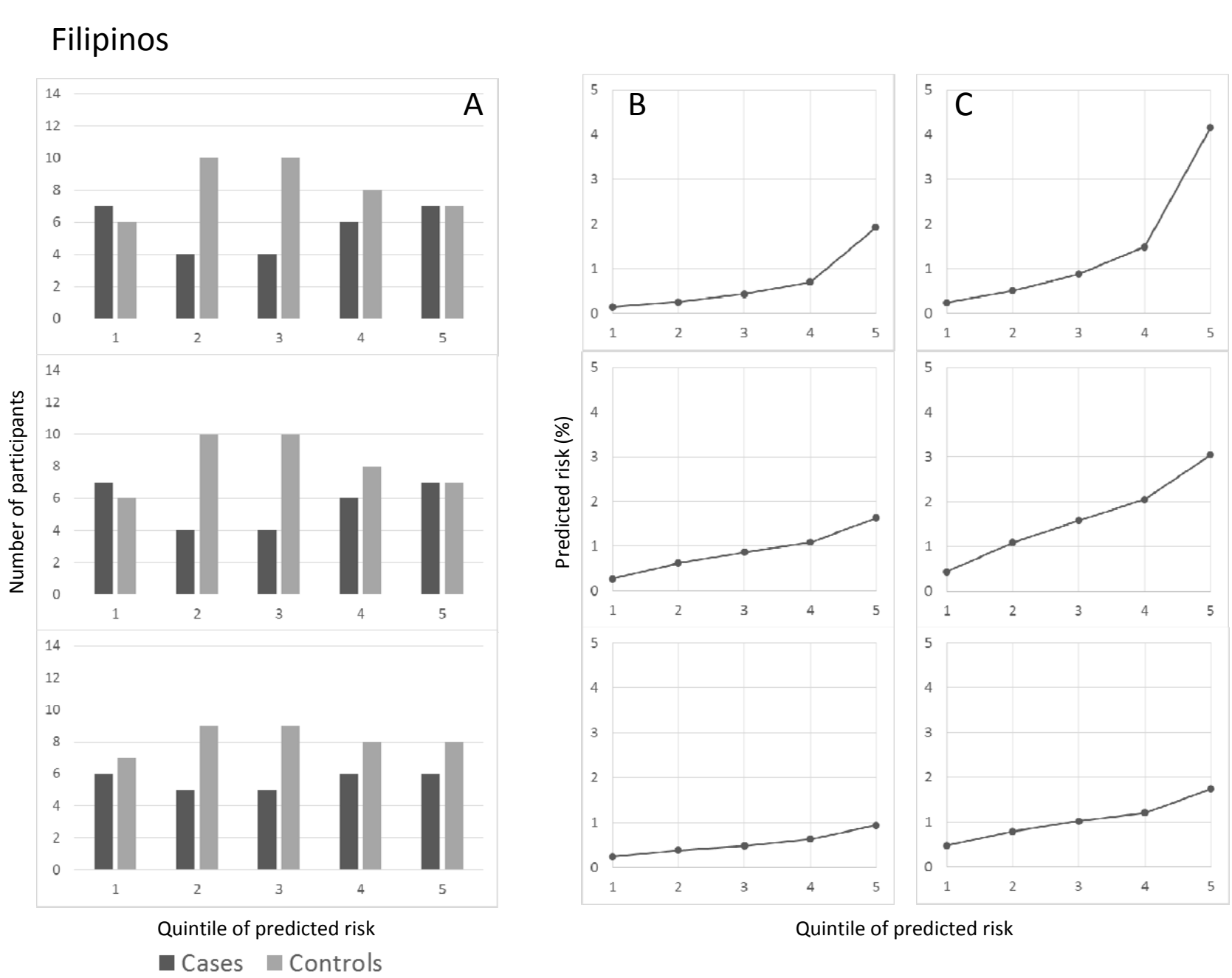
32. Pinhey TK, Heathcote GM, Rarick J. The Influence of Obesity on the Self-Reported Health Status of Chamorros and other Residents of Guam. Asian Am Pac Isl J Health. 1994;2(3):195-211.
33. Bandera EV, Maskarinec G, Romieu I, *et al.* Racial and Ethnic Disparities in the Impact of Obesity on Breast Cancer Risk and Survival: A Global Perspective. Adv Nutr. 2015;6(6):803-19.
34. Nagrani R, Mhatre S, Rajaraman P, *et al.* Central obesity increases risk of breast cancer irrespective of menopausal and hormonal receptor status in women of South Asian Ethnicity. Eur J Cancer. 2016;66:153-61.
35. White KK, Park SY, Kolonel LN, *et al.* Body size and breast cancer risk: the Multiethnic Cohort. Int J Cancer. 2012;131(5):E705-16.
36. Chen J, Pee D, Ayyagari R, *et al.* Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. J Natl Cancer Inst. 2006;98(17):1215-26.
37. Maas P, Barrdahl M, Joshi AD, *et al.* Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA Oncol. 2016;2(10):1295-1302.

Figure 1. Cumulative incidence rates of invasive breast cancer in Guam and the U.S., 2000-2009.



Sources: (1) Guam Cancer Registry; (2) Hawaii Tumor Registry; (3) Surveillance, Epidemiology and End Results (SEER) 18-registry data.





Supplementary Table S1. Invasive breast cancer incidence rates in Guam women, 2000-2009 averages.

Age groups	Rate per 100,000	
	Filipino	Chamorro
20-24	0.0000	0.0000
25-29	6.8880	0.0000
30-34	19.8226	25.5028
35-39	87.2228	24.0964
40-44	76.7575	90.3465
45-49	137.0822	120.2289
50-54	110.8134	202.0321
55-59	186.8822	309.1133
60-64	85.5405	235.3095
65-69	139.3241	337.4036
70-74	227.2385	271.7201
75-79	123.1227	464.6494
80-84	108.3348	390.1445
>85	307.4785	564.2822

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *case-control studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(a) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	6
		(*b*) For matched studies, give matching criteria and the number of controls per case	6-7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-8
Bias	9	Describe any efforts to address potential sources of bias	6
Study size	10	Explain how the study size was arrived at	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-10
		(*b*) Describe any methods used to examine subgroups and interactions	8-10
		(*c*) Explain how missing data were addressed	8
		(*d*) If applicable, explain how matching of cases and controls was addressed	8
		(*e*) Describe any sensitivity analyses	9-10
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	7
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10-12
		(b) Indicate number of participants with missing data for each variable of interest	12

Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure	10-12
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	12-13
		(*b*) Report category boundaries when continuous variables were categorized	10-11
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	12
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12-13
Discussion			
Key results	18	Summarise key results with reference to study objectives	14
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	16
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	16
Generalisability	21	Discuss the generalisability (external validity) of the study results	14, 16
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17-18

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

TRIPOD Checklist: Prediction Model Development

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2-3
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	4-5
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	5
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	6-7
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	6
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	6
	5b	Describe eligibility criteria for participants.	6
	5c	Give details of treatments received, if relevant.	N/A
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	5, 6
	6b	Report any actions to blind assessment of the outcome to be predicted.	N/A
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	6-7
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	Explain how the study size was arrived at.	7
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	8
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	8-9
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	8-9
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	9-10
Risk groups	11	Provide details on how risk groups were created, if done.	9
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	10
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	10-11
Model development	14a	Specify the number of participants and outcome events in each analysis.	10
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	12-13
	15b	Explain how to the use the prediction model.	9, 11
Model performance	16	Report performance measures (with CIs) for the prediction model.	12-13
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	16-17
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	17
Implications	20	Discuss the potential clinical use of the model and implications for future research.	15-16
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	18
Funding	22	Give the source of funding and the role of the funders for the present study.	18

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.