

ARTICLE

Projecting Individualized Absolute Invasive Breast Cancer Risk in US Hispanic Women

Matthew P. Banegas, Esther M. John, Martha L. Slattery, Scarlett Lin Gomez, Mandi Yu, Andrea Z. LaCroix, David Pee, Rowan T. Chlebowski, Lisa M. Hines, Cynthia A. Thompson, Mitchell H. Gail

Affiliations of authors: The Center for Health Research, Kaiser Permanente Northwest, Portland, OR (MPB); Cancer Prevention Institute of California, Fremont, CA (EMJ, SLG); Department of Health Research and Policy (Epidemiology) and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA (EMJ, SLG); University of Utah, Department of Medicine, Salt Lake City, UT (MLS); Surveillance Research Program, Division of Cancer Control and Population Sciences (MY) and Division of Cancer Epidemiology and Genetics (MHG), Mitchell H. Gail (MHG), National Cancer Institute, Bethesda, MD (MY); Division of Epidemiology, Family and Preventive Medicine, University of California, San Diego, La Jolla, CA (AZL); Information Management Services, Rockville, MD (DP); Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA (RTC); Department of Biology, University of Colorado Colorado Springs, Colorado Springs, CO (LMH); Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ (CAT)

Correspondence to: Matthew P. Banegas, PhD, MPH, Center for Health Research, Kaiser Permanente Northwest, 3800 N. Interstate Avenue, Portland, OR 97227-1110 (e-mail: matthew.p.banegas@kpchr.org).

Abstract

Background: There is no model to estimate absolute invasive breast cancer risk for Hispanic women.

Methods: The San Francisco Bay Area Breast Cancer Study (SFBCS) provided data on Hispanic breast cancer case patients (533 US-born, 553 foreign-born) and control participants (464 US-born, 947 foreign-born). These data yielded estimates of relative risk (RR) and attributable risk (AR) separately for US-born and foreign-born women. Nativity-specific absolute risks were estimated by combining RR and AR information with nativity-specific invasive breast cancer incidence and competing mortality rates from the California Cancer Registry and Surveillance, Epidemiology, and End Results program to develop the Hispanic risk model (HRM). In independent data, we assessed model calibration through observed/expected (O/E) ratios, and we estimated discriminatory accuracy with the area under the receiver operating characteristic curve (AUC) statistic.

Results: The US-born HRM included age at first full-term pregnancy, biopsy for benign breast disease, and family history of breast cancer; the foreign-born HRM also included age at menarche. The HRM estimated lower risks than the National Cancer Institute's Breast Cancer Risk Assessment Tool (BCRAT) for US-born Hispanic women, but higher risks in foreign-born women. In independent data from the Women's Health Initiative, the HRM was well calibrated for US-born women (observed/expected [O/E] ratio = 1.07, 95% confidence interval [CI] = 0.81 to 1.40), but seemed to overestimate risk in foreign-born women (O/E ratio = 0.66, 95% CI = 0.41 to 1.07). The AUC was 0.564 (95% CI = 0.485 to 0.644) for US-born and 0.625 (95% CI = 0.487 to 0.764) for foreign-born women.

Conclusions: The HRM is the first absolute risk model that is based entirely on data specific to Hispanic women by nativity. Further studies in Hispanic women are warranted to evaluate its validity.

The National Cancer Institute's (NCI's) Breast Cancer Risk Assessment Tool (BCRAT) predicts invasive breast cancer risk in non-Hispanic white (NHW) (see "model 2" in [1]), African American (2), and Asian and Pacific Islander American (3) women. BCRAT also estimates risk for Hispanic women by

combining Hispanic age-specific incidence rates from the NCI's Surveillance, Epidemiology, and End Results (SEER) program with relative and attributable risks from white women. In a previous study, BCRAT underestimated breast cancer risk by 18% for Hispanic women, and the relative risk estimates for Hispanics

also differed from those in BCRAT (4). Differences between Hispanic and NHW women in distributions of breast cancer risk factors and their relative risks have been reported (5–9). Furthermore, country of birth statistically significantly modifies breast cancer risk (10). Therefore, we developed a breast cancer risk prediction model for Hispanic women based on nativity-specific data.

The San Francisco Bay Area Breast Cancer Study (SFBCS), a multiethnic population-based case-control study, collected data on Hispanic women (10). We combined relative and attributable risks from SFBCS with nativity-specific Hispanic breast cancer incidence and mortality data from the California Cancer Registry (CCR) and SEER program to build the Hispanic risk model (HRM). This model estimates absolute invasive breast cancer risk separately for Hispanic women born in the United States and foreign-born Hispanic women. We compared risk projections from the HRM with those from BCRAT and assessed calibration with independent data from the 4-Corners Breast Cancer Study (4-CBCS) and the Women's Health Initiative (WHI).

Methods

Data Sources

The SFBCS is a population-based case-control study of breast cancer in Hispanic, African American, and NHW women residing in the San Francisco Bay Area. Case patients age 35 to 79 years were diagnosed with a first primary invasive breast cancer between April 1995 and April 2002 (10). Control participants were identified by random-digit dialing and frequency-matched to case patients on five-year age group and race/ethnicity. Race/ethnicity was based on self-report. Non-Hispanic women were excluded. Complete model risk factor data were available for 1086 Hispanic case patients (533 US-born, 553 foreign-born) and 1411 Hispanic control participants (464 US-born, 947 foreign-born); 12 case patients and 29 control participants with missing data were excluded.

The 4-CBCS, a population-based case-control study of breast cancer, included Hispanic, Native American, and NHW women residing in non-reservation areas of Arizona, Colorado, New Mexico, and Utah. Case patients age 25 to 79 years were diagnosed with histologically confirmed in situ or invasive breast cancer between October 1999 and May 2004 (11). Control participants were frequency-matched to case patients based on five-year age intervals and ethnicity. Women who were non-Hispanic or diagnosed with in situ breast cancer or second primary breast cancer were excluded. Complete model risk factor data were available for 592 Hispanic case patients and 802 Hispanic control participants 166 case patients and 81 participants with missing data were excluded.

Both the SFBCS and 4-CBCS collected information on age at diagnosis case patients or selection into the study control participants race/ethnicity, family history of breast cancer in first-degree female relatives, age at menarche, parity, and age at first full-term pregnancy. As part of the Breast Cancer Health Disparities Study, the data from the two studies were harmonized using common definitions (12). For the current analysis, women with a history of ovarian breast cancer were excluded. Information on participants' nativity and previous biopsy for benign breast disease were available from the SFBCS only. We therefore used the SFBCS for model development.

CCR and SEER Data

We obtained age- and nativity-specific (US-born vs foreign-born) invasive breast cancer incidence rates for Hispanic

women from an enhanced CCR data set provided by the Cancer Prevention Institute of California, and competing mortality rates from SEER. CCR data included all Hispanic female California residents age 25 years or older and diagnosed with first primary invasive breast cancer between 1995 and 2004. SEER mortality data included all Hispanic female California residents age 25 years or older at diagnosis who died from non-breast cancer causes between 1995 and 2004. Hispanic nativity was classified as previously described and validated (13).

Women's Health Initiative

The WHI is a national, longitudinal study composed of a set of randomized clinical trials and an observational study (14–16). We used data on 6220 postmenopausal Hispanic women age 50 to 79 years who entered the WHI study without a history of breast cancer or mastectomy (bilateral or unilateral) and who were followed through March 2005 (WHI Main Study). We excluded women with missing information on previous breast cancer or mastectomy.

Statistical Analysis

To select risk factors, we examined variables in SFBCS that most closely aligned with the five independent risk factors and coding in the Gail model (17), but also considered other risk factors (see the [Supplementary Methods, available online](#)). Following the previously described methods (2,3,17), we first developed nativity-specific multivariable relative risk models from the SFBCS data, including risk factors in (17). Then, we obtained baseline age- and nativity-specific breast cancer incidence rates for Hispanic women by multiplying age- and nativity-specific rates from CCR times 1 minus the nativity-specific population-attributable risk from SFBCS. Finally, we calculated absolute risk projections for a Hispanic woman with a specific risk factor profile by multiplying her multivariable relative risk times the baseline age-specific breast cancer incidence rate and accounting for competing mortality rates from SEER, as described below.

Nativity-specific odds ratios were obtained using conditional logistic regression, conditioning on five-year age groups, separately for US-born and foreign-born Hispanic women in SFBCS, with independent variables shown in [Table 1](#). Specifically, the log-relative odds main effects model included the following risk factors as of the reference year (ie, the calendar year before diagnosis for case patients or before selection into the study for control participants age at first full-term pregnancy ($0 = <20$ years, $1 = 20$ – 29 years, $2 = \geq 30$ years or nulliparous); age at menarche ($2 = <12$ years, $1 = 12$ – 13 years, and $0 = \geq 14$ years); history of first-degree relatives with breast cancer ($0 =$ no relatives, $1 = \geq 1$ relatives); history of biopsy with benign breast disease ($0 =$ no biopsy, $1 = \geq 1$ biopsies). Age (in five-year age categories) was included to account for matching. The value of the log odds corresponding to variables in [Table 1](#) and their estimated variance-covariance matrix are shown in [Supplementary Table 1 \(available online\)](#). The [Supplementary Methods \(available online\)](#) give additional details on coding and variable selection.

We calculated the conversion factor $F(t) = 1 - AR(t)$, where $AR(t)$ is the population-attributable risk at age t , from (18):

$$1 - AR(t) = \Sigma(1/rr),$$

where the sum of reciprocal estimated relative risks, $1/rr$, is over the case patients age t with complete data in SFBCS. This formula was applied separately for case patients in four groups defined by ages ($t < 50$ years vs $t \geq 50$ years) and by nativity.

Table 1. Multivariable relative risk estimates for US-born and foreign-born Hispanic women from the San Francisco Area Breast Cancer Study*

US-born Hispanics Risk factor (assigned code)	RR (95% CI)	No. of case patients (n = 533)	No. of control patients (n = 464)
Age at first full-term pregnancy (AFP), y			
<20 (0)	1.0 (Referent)	124	134
20–29 (1)	1.26 (1.05 to 1.52)	280	242
≥30/nulliparous (2)	1.59 (1.10 to 2.31)	129	88
Biopsy for benign breast disease (BIOP)			
No (0)	1.0 (Referent)	431	385
Yes (1)	1.10 (0.79 to 1.53)	102	79
Family history of breast cancer in first-degree female relatives (FH)			
No (0)	1.0 (Referent)	445	396
Yes (1)	1.18 (0.83 to 1.68)	88	68
Foreign-born Hispanics Risk factor (assigned code)	RR (95% CI)	No. of case patients (n = 553)	No. of control patients (n = 947)
Age at first full-term pregnancy (AFP), y			
<20 (0)	1.0 (Referent)	100	267
20–29 (1)	1.60 (1.35 to 1.88)	307	526
≥30/nulliparous (2)	2.54 (1.84 to 3.53)	146	154
Biopsy for benign breast disease (BIOP)			
No (0)	1.0 (Referent)	468	856
Yes (1)	1.62 (1.16 to 2.24)	85	49
Family history of breast cancer in first-degree female relatives (FH)			
No (0)	1.0 (Referent)	486	898
Yes (1)	2.48 (1.67 to 3.68)	85	49
Age at menarche (MEN), y			
≥14 (0)	1.0 (Referent)	165	370
12–13 (1)	1.30 (1.12 to 1.50)	248	391
<12 (2)	1.68 (1.26 to 2.25)	140	186

*Estimates obtained from multivariable conditional logistic regression models separately for US-born and foreign-born Hispanic women, with risk factors coded as shown. The overall relative risk of developing invasive breast cancer, compared with an individual of the same nativity with all risk factors at their lowest risk level, is estimated by multiplying the relative risks across risk factors. CI = confidence interval; RR = relative risk.

We used age- and nativity-specific invasive breast cancer incidence rates $h^*(t)$ in five-year age intervals from the CCR (Supplementary Table 2, available online) to estimate the baseline hazard as $h_1(t) = h^*(t)F(t)$. The hazard $h_2(t)$ of age- and nativity-specific mortality from non-breast cancer causes was obtained from SEER data (Supplementary Table 3, available online). Using Equation 6 in (17) with one-year intervals, we combined the information on h_1 , h_2 , and the relative risk (rr) to project individualized absolute risk of invasive breast cancer for various initial and final ages and combinations of risk factors.

We estimated the variance of the absolute risk by bootstrapping SFBCS data. From 10 000 bootstrap samples, we obtained 10 000 estimates of the log-relative risks and attributable risk. For a fixed combination of risk factors and risk projection interval, we obtained 10 000 bootstrap estimates of absolute risk by regarding the h^* and h_2 as known quantities, but applying the 10 000 different sets of relative risks and attributable risk. The 95% confidence interval (CI) was defined by the 2.5th percentile and 97.5th percentile of the resulting bootstrap distribution of absolute risk. We refer to these confidence limits for a given risk projection as “exact,” and we developed a SAS computer program to compute them. We also developed simple graphs (Figure 1, A and B, respectively, for US-born and foreign-born Hispanic women) that plot approximate confidence limits against projected absolute risk. Thus, one can obtain approximate confidence intervals by estimating absolute risk for any combination of risk factors and projection interval and

reading the approximate confidence limits from Figure 1. The methods to construct Figure 1 are in the Supplementary Methods (available online).

Projections for HRM with confidence intervals can be computed using a SAS macro (HisBrCa_RAM_Ver2_1, available at http://dceg.cancer.gov/tools/risk-assessment/HisBrCa_RAM/). This program was developed using SAS 9.3 (SAS Institute Inc., Cary, NC).

To assess the relative risks in the HRM, we estimated them from independent data in 4-CBCS and WHI. In 4-CBCS, we used multivariable logistic regression to estimate the odds ratios for age at first full-term pregnancy, age at menarche, and family history of breast cancer in first-degree female relatives, regardless of nativity. In WHI, we used Cox proportional hazards models to estimate hazard ratios for age at first live birth, age at menarche, family history of breast cancer in first-degree relatives, and biopsy for benign breast disease, separately by nativity. Because we wanted to compare relative risks from WHI with those from HRM, we used the same models for covariates. We tested the proportionality assumption using scaled Schoenfeld residuals. We assessed the calibration and discriminatory accuracy of the HRM in WHI as in (4) by computing each woman's absolute risk of developing invasive breast cancer from WHI enrollment through March 2005. The expected counts in risk factor category i , E_i , were calculated as the sum of risks of women in that category and compared with the observed number of

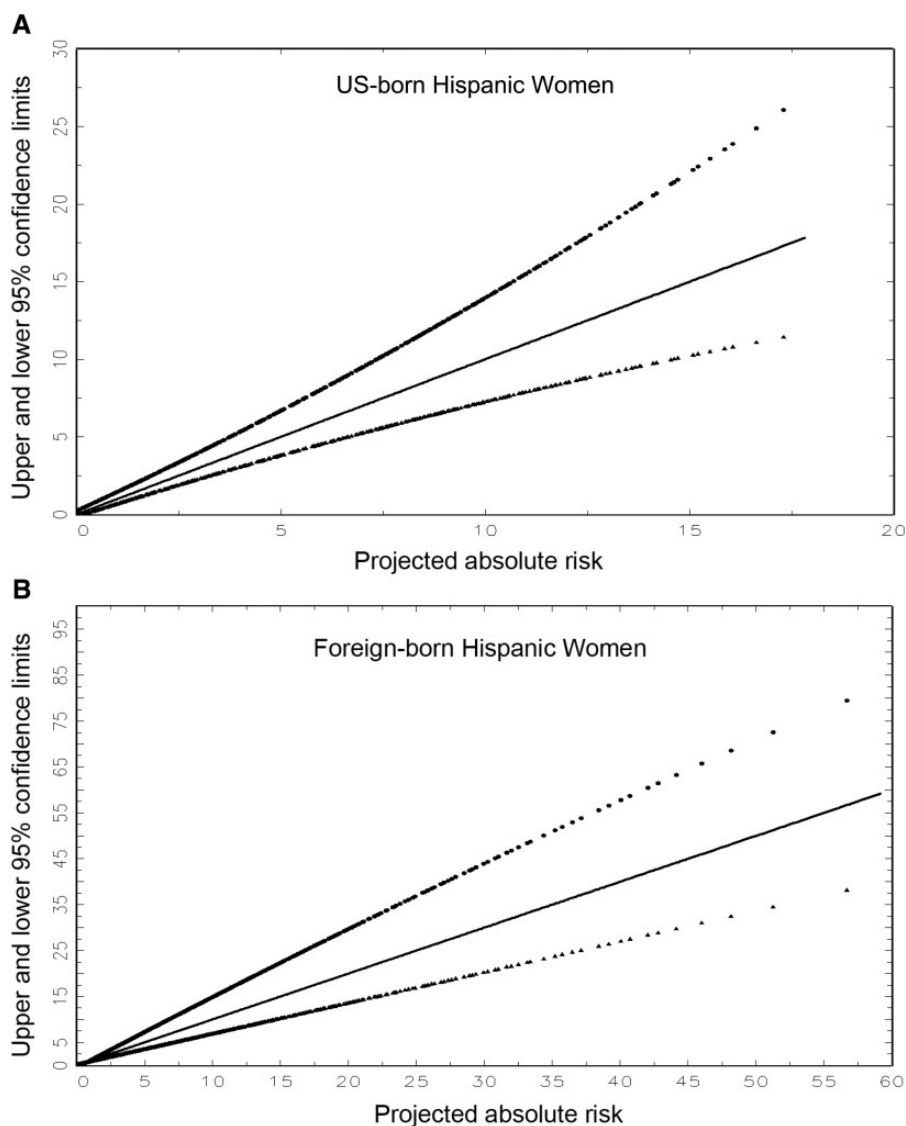


Figure 1. Approximate upper and lower 95% confidence limits for the estimated absolute risk of invasive breast cancer plotted against the projected absolute risk. **A)** For US-born Hispanic women. **B)** For foreign-born Hispanic women. Confidence limits were calculated as described in the “Analytic Approach and Risk Factors” section. The loci represent regressions that are quadratics in absolute risk.

women with incident invasive breast cancer, O_i . For each category, we computed an observed/expected (O/E) ratio and 95% confidence interval from $(O/E)\exp(\pm 1.96 \times O^{-1/2})$. Discriminatory accuracy comparing WHI patients with non-case patients was assessed using the concordance statistic, or area under the receiver operating characteristic curve (AUC) statistic (19). Statistical significance tests were two-sided, with the significance level set at a P value of less than .05.

Results

Relative and Attributable Risks

Relative risks for the HRM are given separately for US-born and foreign-born Hispanic women (Table 1). The corresponding conversion factors $F(t)=1-AR(t)$ were 0.749 (95% CI = 0.601 to 0.926) for US-born women younger than age 50 years, 0.778 (95% CI = 0.613 to 0.890) for US-born women age 50 years or older, 0.429 (95% CI = 0.336 to 0.537) for foreign-born women younger than

age 50 years, and 0.450 (95% CI = 0.350 to 0.514) for foreign-born women age 50 years or older.

Individualized Absolute Risks

Table 2 provides absolute risk estimates for various initial ages, follow-up durations, and relative risks for US-born and foreign-born Hispanic women. Suppose one wishes to project invasive breast cancer risk over 20 years for a 30-year-old foreign-born Hispanic woman who had her first full-term pregnancy at age 28 years (AFP = 1), who never had a biopsy for benign breast disease (BIOP = 0), whose mother had breast cancer (FH = 1), and who began menstruating at age 11 years (MEN = 2). From Table 1, the woman’s relative risk is 1.60 (for AFP = 1) x 1.00 (for BIOP = 0) x 2.48 (for FH = 1) x 1.68 (for MEN = 2) = 6.67. From Table 2, the woman’s 20-year absolute risk is between 2.56% for relative risk 5 and 5.06% for relative risk 10. By linear interpolation, the approximate risk is $2.56 + (5.06 - 2.56)(6.67 - 5.00)/(10 - 5) = 3.40\%$. The exact estimate from our SAS program, `HispBrCa_RAM_Ver2_1`, is 3.39%.

Table 2. Projected absolute risk (%) of developing invasive breast cancer within 5, 10, 20, or 30 years by relative risk, initial age, and years of follow-up for Hispanic women*

Initial age, y	Years of follow-up	Project absolute risk, %							
		US-born Hispanic initial relative risk				Foreign-born Hispanic initial relative risk			
		1	2	5	10	1	2	5	10
20	5	0.01	0.01	0.03	0.06	0.00	0.00	0.01	0.02
	10	0.03	0.07	0.17	0.34	0.01	0.03	0.07	0.14
	20	0.36	0.72	1.80	3.56	0.12	0.25	0.62	1.23
	30	1.50	2.98	7.27	14.01	0.53	1.06	2.62	5.17
30	5	0.10	0.20	0.51	1.02	0.03	0.07	0.17	0.34
	10	0.33	0.66	1.64	3.25	0.11	0.22	0.55	1.10
	20	1.47	2.92	7.14	13.77	0.52	1.03	2.56	5.06
	30	3.34	6.57	15.62	28.75	1.20	2.38	5.85	11.35
40	5	0.46	0.91	2.26	4.46	0.16	0.33	0.81	1.62
	10	1.15	2.29	5.64	10.95	0.41	0.82	2.03	4.02
	20	3.05	6.00	14.32	26.55	1.09	2.18	5.35	10.41
	30	5.14	10.00	23.09	40.69	2.00	3.97	9.62	18.30
50	5	0.89	1.77	4.37	8.55	0.32	0.63	1.57	3.11
	10	1.95	3.86	9.36	17.84	0.69	1.38	3.43	6.73
	20	4.10	8.02	18.83	34.02	1.62	3.21	7.83	15.03
	30	6.07	11.74	26.68	45.85	2.59	5.11	12.27	22.96
60	5	1.18	2.35	5.77	11.20	0.46	0.91	2.26	4.47
	10	2.28	4.51	10.88	20.55	0.96	1.90	4.69	9.16
	20	4.37	8.53	19.90	35.58	1.96	3.89	9.42	17.90
70	5	1.23	2.44	5.99	11.61	0.55	1.10	2.73	5.39
	10	2.36	4.66	11.22	21.10	1.10	2.19	5.37	10.44

*Tables present absolute risk estimates for various initial ages, follow-up durations, and relative risks for US-born and foreign-born Hispanic women. For example, to project invasive breast cancer risk over 20 years for a 30-year-old foreign-born Hispanic woman who had her first full-term pregnancy at age 28 years (age at first full-term pregnancy [AFP] = 1), who never had a biopsy for benign breast disease (biopsy for benign breast disease [BIOP] = 0), whose mother had breast cancer (family history of breast cancer in first-degree female relatives [FH] = 1), and who began menstruating at age 11 years (age at menarche [MEN] = 2). From Table 1, the woman's relative risk is 1.60 (for AFP = 1) x 1.00 (for BIOP = 0) x 2.48 (for FH = 1) x 1.68 (for MEN = 2) = 6.67. As shown in the table, the woman's 20-year absolute risk is between 2.56% for relative risk 5 and 5.06% for relative risk 10. By linear interpolation, the approximate risk is $2.56 + (5.06 - 2.56)(6.67 - 5.00)/(10 - 5) = 3.40\%$.

Confidence Intervals on Risk Projections

For this example with a risk of 3.39%, the approximate 95% confidence interval (Figure 1) was 2.45% to 4.80%, which agrees well with the "exact" bootstrap 95% confidence interval (2.31% to 5.09%). For most purposes, Figure 1 yields sufficiently accurate confidence intervals.

Comparison With NCI's BCRAT

We plotted five-year absolute risks from the HRM (ordinate) against absolute risks from BCRAT (abscissa) separately for US-born Hispanic women age 35, 50, and 70 years (Figure 2, A–C) and for foreign-born Hispanic women age 35, 50, and 70 years (Figure 2, D–F). For US-born women, BCRAT yielded higher absolute risk estimates than the HRM for most risk factor patterns (points below the equiangular line). The percentages of risk patterns in which the BCRAT had higher risk estimates than the HRM were 76%, 55%, and 69%, respectively, in women age 35, 50, and 70 years. For foreign-born women, absolute risks from BCRAT were higher than from the HRM for women

age 35 years in 61% of risk patterns, but lower in 66% for women age 50 years and in 66% for women age 70 years.

Hispanic Risk Model Validation

We compared relative risks from the HRM with those for Hispanic women in 4-CBCS and WHI (Supplementary Table 4, available on line). The relative risk estimates from 4-CBCS are not strictly comparable with those in the HRM because data on nativity and biopsy for benign breast disease were not available in 4-CBCS. None of the 4-CBCS relative risks were statistically significantly different from the relative risks in the US-born HRM. However, the 4-CBCS relative risks for age at first full-term pregnancy (relative risk [RR] = 1.19, 95% CI = 1.01 to 1.41) and family history (RR = 1.39, 95% CI = 1.03 to 1.88) were statistically significantly lower than the respective relative risks (RR = 1.60, 95% CI = 1.35 to 1.88, $P = .01$; RR = 2.48, 95% CI = 1.67 to 3.68, $P = .02$) in the foreign-born HRM. In WHI, there were too few invasive breast cancer events to fit reliable multivariable relative risk models. For foreign-born WHI Hispanic women, none of the relative risks were statistically different from those in the foreign-born HRM; whereas for US-born

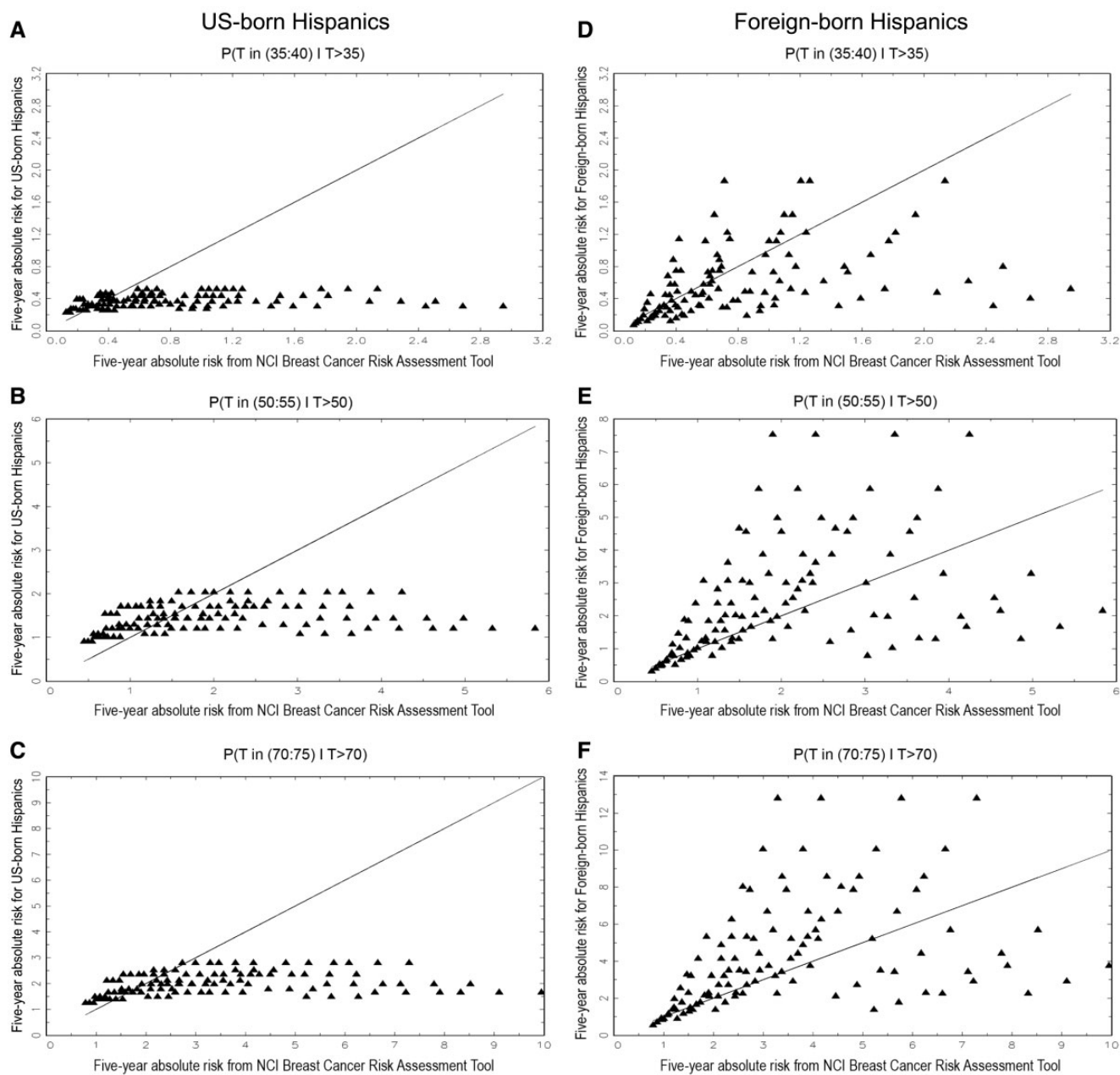


Figure 2. Plots of five-year projections of absolute invasive breast cancer risk in Hispanic women based on the HRM (ordinate) vs the NCI Breast Cancer Risk Assessment Tool (BCRAT) (abscissa). **A)** For US-born Hispanic women age 35 years. **B)** For US-born Hispanic women age 50 years. **C)** For US-born Hispanic women age 70 years. **D)** For foreign-born Hispanic women age 35 years. **E)** For foreign-born Hispanic women age 50 years. **F)** For foreign-born Hispanic women age 70 years.

WHI Hispanic women, the relative risk for age at first full-term pregnancy (RR = 0.71, 95% CI = 0.47 to 1.08) was statistically significantly lower than that in the US-born HRM (RR = 1.26, 95% CI = 1.05 to 1.52, $P = .01$).

We assessed the HRM calibration and discriminatory accuracy in WHI (Table 3). For US-born Hispanic women ($n = 2079$), 52 breast cancers were diagnosed (“observed”), compared with 48.7 expected from the HRM, yielding an O/E of 1.07 (95% CI = 0.81 to 1.40). Thus, the US-born HRM was well calibrated in WHI. For foreign-born women ($n = 1230$), 17 breast cancers were observed compared with 25.6 expected, yielding an O/E of 0.66 (95% CI = 0.41 to 1.07). Thus, the HRM appears to overestimate risk in this

group, but the deviation was not statistically significant. For WHI women of unknown nativity ($n = 2911$), 61 breast cancers were observed and 68.7 were expected, yielding an O/E of 0.89 (95% CI = 0.69 to 1.14). [Supplementary Table 5 \(available online\)](#) gives data on O and E within groups defined by tertiles of risk. There is no evidence of overfitting as the agreement between O and E is similar across tertile categories. The nativity-specific concordance statistics (AUC) of the HRM in the WHI cohort (Table 3) were an AUC of 0.564 (95% CI = 0.485 to 0.644) for US-born Hispanic women, an AUC of 0.625 (95% CI = 0.487 to 0.764) for foreign-born Hispanic women, and an AUC of 0.582 (95% CI = 0.509 to 0.656) for Hispanic women of unknown nativity.

Table 3. Analysis of observed vs expected numbers of invasive breast cancers among Hispanic women in the Women's Health Initiative*

Risk category	Hispanic risk model				
	AUC (95% CI)	No. of women	No. observed	No. expected	O/E ratio (95% CI)
US-born Hispanics	0.564 (95% CI = 0.485 to 0.644)	2079	52	48.73	1.07 (0.81 to 1.40)
Foreign-born Hispanics	0.625 (95% CI = 0.487 to 0.764)	1230	17	25.59	0.66 (0.41 to 1.07)
Hispanics of unknown nativity	0.582 (95% CI = 0.509 to 0.656)	2911	61	68.71	0.89 (0.69 to 1.14)

*Counts represent the number of women in each nativity group and the number of incident invasive breast cancers observed during the Women's Health Initiative (WHI) Main Study period. Expected cancers are calculated by applying the Hispanic risk model (HRM) to Hispanic women in WHI. The WHI did not collect nativity information for all participants; specifically, women who participated in the clinical trial arm did not have nativity information and are categorized as "unknown nativity" for this analysis. For Hispanic WHI women of unknown nativity, we calculated the expected count as a weighted average of the expected counts from the US-born and foreign-born HRMs, based on the proportion of US-born women among women with known nativity, $2079/(2079 + 1230) = 0.6283$. AUC = area under the receiver operating characteristic curve; CI = confidence interval; O/E=observed-to-expected ratio.

Discussion

We developed a nativity-specific model to project individualized, absolute invasive breast cancer risk for US Hispanic women. A SAS program provides "exact" risk estimates with confidence intervals, but [Tables 1 and 2](#) can be used to obtain nativity-specific risk projections, and corresponding approximate confidence intervals can be read from [Figure 1](#). Hispanics are the largest racial/ethnic minority group in the United States, comprising nearly 17% of the total US population (approximately 54 million in 2013), with US-born Hispanics representing the largest share (65%) (20).

To incorporate nativity, we obtained nativity-specific incidence and mortality rates and used nativity-specific relative risks from SFBCS. We omitted some risk factors and simplified the coding of some risk factors, compared with the original Gail model (see [Supplementary Methods, available online](#)) (17). The resulting simplified HRM model fit the SFBCS data well and yielded absolute risk estimates with smaller variances than with the original coding. Projections of five-year absolute risk from the HRM were usually lower than those from the NCI BCRAT for US-born Hispanic women, but were usually higher for foreign-born Hispanic women age 50 years or older.

When assessing the HRM with independent data, relative risk estimates from 4-CBCS were similar to those for the US-born HRM, but lower than the relative risks for age at first full-term pregnancy and family history in the foreign-born HRM. The US-born HRM was well calibrated in WHI US-born Hispanic women. In foreign-born WHI women, the HRM appeared to overestimate risk, but this discrepancy was not statistically significant. In WHI, the concordance statistics for the HRM for US-born and foreign-born Hispanic women, while modest, were similar to those obtained for other BCRAT models (2,3).

Because the HRM was based on data from SFBCS, CCR, and SEER on all Hispanic female California residents, the HRM is likely to be appropriate for Hispanic women with origins similar to those in the SFBCS and CCR—namely, Hispanic women in western US states, primarily of Mexican and Central American descent. Studies highlighting heterogeneity in breast cancer risk between Hispanic women underscore the importance of differences in country of origin, duration of residence in the United States, and acculturation in estimating the risk of breast cancer (10,21–23). Further, there is evidence of genomic differences between Hispanic subgroups in the United States (24,25). These factors were not captured in the HRM and may contribute to some of the differences in HRM relative risks and absolute risk predictions in the validation studies. Future studies are warranted that collect comprehensive information on breast cancer

risk factors, genomic data, and health outcomes across different populations of Hispanic women. Such data can be used to evaluate the HRM further and to develop improved risk prediction models.

We note several additional limitations. First, the HRM had modest discriminatory accuracy, which highlights the need for considering additional risk factors, such as mammographic density (26–29) or genetic variants (30,31). Nevertheless, an advantage of the HRM, like BCRAT, is that the information required is available from self-report. Second, the validation data sets did not have data on all HRM risk factors (eg, nativity in 4-CBCS) or covered restricted age ranges (eg, participant age was ≥ 50 years only in WHI), limiting our ability to test calibration across risk factor categories. However, these studies are among the few sources with information on breast cancer risk factors in Hispanic women. Third, in addition to random error in projections, there is the possibility of bias from misspecification of the model. Fourth, the ages of participants in SFBCS were 35 to 79 years; projections outside these ages assumed constancy of relative and attributable risks. Further efforts to evaluate the HRM will require more extensive data on Hispanic women.

The HRM, like the BCRAT, should not be used for certain women. It will probably underestimate breast cancer risk in Hispanic women with a personal history of invasive breast cancer or ductal or lobular carcinoma in situ, in women carrying breast cancer-causing mutations (eg, *BRCA1* or *BRCA2*), and in women who received therapeutic radiation doses to the breast at a young age, such as for treatment of Hodgkin's lymphoma (32).

Funding

This work was supported by the Cancer Prevention Fellowship Program of the Division of Cancer Prevention and the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health. The San Francisco Bay Area Breast Cancer Study (SFBCS) was supported by grants CA63446 and CA77305 (to E. M. John) from the National Cancer Institute, grant DAMD17-96-1-6071 (to E. M. John) from the US Department of Defense, and grant 7PB-0068 (to E. M. John) from the California Breast Cancer Research Program. The 4-Corners Breast Cancer Study was funded by grants CA078682, CA078762, CA078552, and CA078802 from the National Cancer Institute. The Breast Cancer Health Disparities Study was funded by grant CA14002 (to M. L. Slattery) from the National Cancer Institute. Dr. Gomez was supported by the National

Cancer Institute's Surveillance, Epidemiology, and End Results Program under contract NNSH261201000140C awarded to the Cancer Prevention Institute of California and by the Stanford Cancer Institute. The collection of cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance,

Epidemiology, and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement U58DP003862-01 awarded to the California Department of Public Health. The Women's Health Initiative (WHI) program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, and US Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN26820110004C, and HHSN271201100004C.

Notes

The study funders had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication. The ideas and opinions expressed herein are those of the authors, and endorsement by the State of California, the Department of Public Health, the National Cancer Institute, or the Centers for Disease Control and Prevention or their contractors and subcontractors is not intended and should not be inferred.

We would also like to acknowledge the contributions of Dr. Hormuzd Katki for study development, Dr. Dorothy Lane for input on the manuscript, Dr. Jean Wactawski-Wende for continued support of this project with the WHI, Jennifer Herrick for data management and data harmonization of the Breast Cancer Health Disparities Study; Jocelyn Koo for data management for the San Francisco Bay Area Breast Cancer Study; and Drs. Kathy Baumgartner, Tim Byers, and Anna Giuliano for their contributions to the 4-Corners Breast Cancer Study.

References

- 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):1541–1548.
- 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):1782–1792.
- 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):951–961.
- Banegas MP, Gail MH, Lacroix A, et al. Evaluating breast cancer risk projections for Hispanic women. *Breast Cancer Res Treat.* 2012;132(1):347–353.
- Hines LM, Risendal B, Slattery ML, et al. Comparative analysis of breast cancer risk factors among Hispanic and non-Hispanic white women. *Cancer.* 2010;116(13):3215–3223.
- Sweeney C, Baumgartner KB, Byers T, et al. Reproductive history in relation to breast cancer risk among Hispanic and non-Hispanic white women. *Cancer Causes Control.* 2008;19(4):391–401.
- Bondy ML, Spitz MR, Halabi S, Fueger JJ, Vogel VG. Low incidence of familial breast cancer among Hispanic women. *Cancer Causes Control.* 1992;3(4): 377–382.
- Risendal B, Hines LM, Sweeney C, et al. Family history and age at onset of breast cancer in Hispanic and non-Hispanic white women. *Cancer Causes Control.* 2008;19(10):1349–1355.
- Chlebowski RT, Chen Z, Anderson GL, et al. Ethnicity and breast cancer: Factors influencing differences in incidence and outcome. *J Natl Cancer Inst.* 2005;97(6):439–448.
- John EM, Phipps AI, Davis A, Koo J. Migration history, acculturation, and breast cancer risk in Hispanic women. *Cancer Epidemiol Biomarkers Prev.* 2005; 14(12):2905–2913.
- Slattery ML, Sweeney C, Edwards S, et al. Body size, weight change, fat distribution and breast cancer risk in Hispanic and non-Hispanic white women. *Breast Cancer Res Treat.* 2007;102(1):85–101.
- Slattery ML, John EM, Torres-Mejia G, et al. Genetic variation in genes involved in hormones, inflammation and energetic factors and breast cancer risk in an admixed population. *Carcinogenesis.* 2012;33(8):1512–1521.
- Keegan TH, John EM, Fish KM, Alfaro-Velcamp T, Clarke CA, Gomez SL. Breast cancer incidence patterns among California Hispanic women: Differences by nativity and residence in an enclave. *Cancer Epidemiol Biomarkers Prev.* 19(5):1208–1218.
- Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials.* 1998; 19(1):61–109.
- Hays J, Hunt JR, Hubbell FA, et al. The Women's Health Initiative recruitment methods and results. *Ann Epidemiol.* 2003;13(9 Suppl):S18–S77.
- Ritenbaugh C, Patterson RE, Chlebowski RT, et al. The Women's Health Initiative Dietary Modification trial: Overview and baseline characteristics of participants. *Ann Epidemiol.* 2003;13(9 Suppl):S87–S97.
- 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):1879–1886.
- Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol.* 1985;122(5):904–914.
- Wieand S, Gail MH, James BR, James KL. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika.* 1989;76(3):585–592.
- Stepler R, Brown A. *Statistical Portrait of the Hispanic Population in the United States, 1980–2013.* Pew Research Center; Washington, DC. 2015.
- Banegas MP, Leng M, Graubard BI, Morales LS. The risk of developing invasive breast cancer in Hispanic women. *Cancer.* 2013;119(7):1373–1380.
- Ooi SL, Martinez ME, Li CI. Disparities in breast cancer characteristics and outcomes by race/ethnicity. *Breast Cancer Res Treat.* 2011;127(3):729–738.
- Keegan TH, John EM, Fish KM, Alfaro-Velcamp T, Clarke CA, Gomez SL. Breast cancer incidence patterns among California Hispanic women: Differences by nativity and residence in an enclave. *Cancer Epidemiol Biomarkers Prev.* 2010; 19(5):1208–1218.
- Conomos MP, Laurie CA, Stilp AM, et al. Genetic diversity and association studies in US Hispanic/Latino populations: Applications in the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet.* 2016;98(1): 165–184.
- Fejerman L, Stern MC, Ziv E, et al. Genetic ancestry modifies the association between genetic risk variants and breast cancer risk among Hispanic and non-Hispanic white women. *Carcinogenesis.* 2013;34(8):1787–1793.
- Vachon CM, van Gils CH, Sellers TA, et al. Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Res.* 2007;9(6):217.
- Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: Development and validation of a new predictive model. *Ann Intern Med.* 2008;148(5):337–347.
- Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treat.* 2005;94(2):115–122.
- 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):1215–1226.
- McCarthy AM, Keller B, Kontos D, et al. The use of the Gail model, body mass index and SNPs to predict breast cancer among women with abnormal (BI-RADS 4) mammograms. *Breast Cancer Res.* 2015;17:1.
- Mealiffe ME, Stokowski RP, Rhee BK, Prentice RL, Pettinger M, Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst.* 2010;102(21):1618–1627.
- Travis LB, Hill D, Dores GM, et al. Cumulative absolute breast cancer risk for young women treated for Hodgkin lymphoma. *J Natl Cancer Inst.* 2005;97(19): 1428–1437.