

**Projecting Individualized Absolute Invasive Breast Cancer Risk in Asian and Pacific
Island American Women**

December 17, 2010

Rayna K. Matsuno, Joseph P. Costantino, Regina G. Ziegler, Garnet L. Anderson, Huilin Li,
David Pee, Mitchell H. Gail

Correspondence to: Mitchell H. Gail, MD, PhD, Division of Cancer Epidemiology and
Genetics, National Cancer Institute, EPS Room 8032, Bethesda, MD 20892-7244 (e-mail:
gailm@mail.nih.gov)

ABSTRACT

Background: The NCI Breast Cancer Risk Assessment Tool (BRCAT) is widely used for estimating absolute risk of invasive breast cancer. However, its estimates for Asian and Pacific Island American (APA) women are based on data from white women. We developed a model for projecting absolute invasive breast cancer risk in APA women and compared its projections to those from BRCAT. *Methods:* Data from the Asian American Breast Cancer Study (AABCS) were used to compute relative and attributable risks based on the age at menarche, number of affected mother or sisters, and number of previous benign biopsies. Absolute risks were obtained by combining this information with ethnicity-specific data from the NCI's Surveillance, Epidemiology and End Results (SEER) Program and with U.S. ethnicity-specific mortality data to create the AABCS model. *Results:* The AABCS model gave absolute risk estimates separately for Chinese, Japanese, Filipino, Hawaiian, Other Pacific Islander, and Other Asian women. Relative risks and attributable risks for APA women were comparable to those in BCRAT, but the AABCS model usually gave lower risk projections than BCRAT in Chinese and Filipino, but not in Hawaiian women and not in every age and ethnic subgroup. The AABCS model underestimated risk by 17% (95% confidence interval -1% to 38%) in independent data from APA women in the Women's Health Initiative, but APA women in the Women's Health Initiative had rates about 18% higher than SEER rates. *Conclusions:* The AABCS model is calibrated to ethnicity-specific SEER rates and is preferable to BCRAT for counseling APA women.

INTRODUCTION

The National Cancer Institute's Breast Cancer Risk Assessment Tool (BCRAT) (<http://www.cancer.gov/bcrisktool/>) projects absolute invasive breast cancer risk and has been used for counseling women and designing breast cancer prevention trials. Although BCRAT includes separate models for white (1) and African American women (2), projections of absolute risk for Asian and Pacific Island American (APA) women are based on data from white women only (1, 3). Therefore, BCRAT includes a disclaimer for APA women. Inaccurate projections could mislead in counseling APA women and might mistakenly render some APA women eligible or ineligible for participation in breast cancer prevention trials. For these reasons, there is a need to develop a model for APA women that is based on sufficient ethnicity-specific data.

The population-based Asian American Breast Cancer Study included 597 Asian American women with invasive breast cancer and 966 Asian American control subjects (4). Because this study gathered information on the factors included in the original Gail model (3), relative and attributable risks specific to APA women could be estimated. In the current study, we used data from the Asian American Breast Cancer Study and ethnicity-specific data from the National Cancer Institute's Surveillance Epidemiology and End Results (SEER) Program from 1998 through 2002 to estimate absolute invasive breast cancer risk for APA women and give 95% confidence intervals for the estimates. We call this new model the Asian American Breast Cancer Study model, or AABCS model. We also compare these new projections with those from the current NCI BCRAT and check the calibration of the new AABCS model with independent data from the Women's Health Initiative (5).

METHODS AND DATA SOURCES

Data sources for constructing the model

The study methods for the population-based Asian American Breast Cancer Study have previously been described in detail in Ziegler et al.(4). Women of Chinese, Japanese, and Filipino ethnicity with histologically confirmed first primary incident breast cancer diagnosed between the ages of 20-55 years were identified through population-based cancer registries in San Francisco-Oakland, California; Los Angeles County, California; and Oahu, Hawaii for the period 1983-1987. All three registries are members of the SEER Program (<http://seer.cancer.gov>). Controls of the same ethnicity, age, and residence were identified through random-digit dialing in the two California areas and through the Hawaii Health Surveillance Program. The final study population consisted of 597 cases (70% of eligible cases) and 966 controls (75% of eligible controls), of whom 589 cases and 952 controls provided complete covariate data for estimating relative risks and attributable risks.

Age- and ethnicity-specific invasive breast cancer incidence rates for Chinese, Japanese, Filipino, Other Asians (excluding the previous three groups), native Hawaiians, and Pacific Islanders (excluding native Hawaiians), were obtained from the SEER Detailed Asian/Pacific Islander Database for the years 1998-2002 (2000-centered) (6). We use the term “ethnicity” to denote these six groups, although the terms “Asian” versus “Hawaiian or other Pacific Islander” have been distinguished as different races (<http://www.whitehouse.gov/omb/rewrite/fedreg/ombdir15.html>). The database represented three metropolitan areas and nine states (Atlanta, Detroit, Seattle/Puget Sound, California, Connecticut, Hawaii, Iowa, Kentucky, Louisiana, New Jersey, New Mexico, and Utah) and thirteen Asian or Pacific Island groups (Chinese, Japanese, Filipino, Asian Indian/Pakistani combined, Korean, Vietnamese, Laotian, Kampuchean, Guamanian, Samoan, Tongan, and

native Hawaiian). These reporting areas covered 54% of the total US Asian and Pacific Islander population and represented 53% of Chinese, 71% of Japanese, and 69% of Filipinos in the U.S. (6). Ethnicities for incident invasive breast cancer cases were obtained from medical records by the SEER cancer registries. Because over 99.95% of cancer diagnoses in SEER include only one ethnicity designation, the SEER Detailed Asian/Pacific Islander Database used only one ethnicity to classify cases (7). The corresponding numbers of women at risk (rate denominators) were based on the U.S. 2000 Census, which allowed individuals to report multiple ethnicities. Therefore, SEER calculated incidence rates using two different methods for determining the number of women at risk. The first method included women who self-reported one ethnicity on the U.S. 2000 Census; the second method included women who self-reported one or more ethnicities, at least one of which was the group of interest. Because the first method results in an overestimate of the true incidence rate and the second method results in an underestimate, we calculated a simple average of the two incidence rates. Unreported calculations indicated that a simple average performs well over a range of (unknown) fractions of women who check multiple ethnicities on a census form but declare themselves to have a specific ethnicity when forced to choose. This procedure was used to calculate rates separately for Chinese, Japanese, Filipino, Other Asians (excluding the previous three groups), native Hawaiians, and Pacific Islanders (excluding native Hawaiians). For native Hawaiians, incidence rates in SEER were calculated using only multiple race/ethnicity denominators, because a case with any native Hawaiian ancestry is classified as native Hawaiian in SEER (6). The resulting age- and ethnicity-specific breast cancer incidence rates are in Supplemental Table 1.

To account for competing risks from non-breast cancer mortality, age- and ethnicity-specific non-breast cancer mortality rates were obtained through SEER from the National Center for Health Statistics (NCHS, <http://www.cdc.gov/nchs>) for the years 1998-2002 (2000-centered) (6) for Chinese, Japanese, Filipino, Other Asians (excluding the previous three groups), native Hawaiians, and Pacific Islanders (excluding native Hawaiians). The database represented seven states (California, Hawaii, Illinois, New Jersey, New York, Texas, and Washington) and nine APA groups (Chinese, Japanese, Filipino, Indian only, Korean, Vietnamese, Guamanian, Samoan, and native Hawaiian). These reporting areas covered 68% of the total US Asian and Pacific Islander population and represented 74% of Chinese, 77% of Japanese, and 79% of Filipinos in the US (6). Ethnicity for non-invasive breast cancer deaths were obtained from state vital records. Because vital records usually include only a single race or ethnicity designation, the NCHS data used only single race or ethnicity information to classify deaths. We calculated the census numbers at risk as previously described for calculating breast cancer incidence rates and used these denominators to calculate mortality rates as previously described for incidence rates (Supplemental Table 1).

Validation data

To assess the calibration of the AABCS model, we used independent data on breast cancer incidence from 4,031 postmenopausal APA women, aged 50-79 who entered the Women's Health Initiative study without a history of breast cancer (5). The women were recruited between 1993 and 1998 and followed for an average of 9.1 years to detect incident invasive breast cancer. Invasive breast cancers were diagnosed at ages ranging from 51.1 to 86.1 years.

Resolving unknown ethnicity category

We imputed the Asian ethnicity for 715 women in the Women's Health Initiative with unknown ethnicity. We used an algorithm developed by the North American Association of Central Cancer Registries (8). The algorithm is based on place of birth, maiden name, surname, or given name, in decreasing order of precedence. When place of birth was unavailable, either maiden name, surname, or given name were checked against the corresponding Census name list (9), the Lauderdale name list (10), or the North American Association of Central Cancer Registries (NAACCR) name list (8) in decreasing order of precedence. After the imputation, 109 women were reclassified as Chinese; 357 were reclassified as Japanese; 87 were reclassified as Filipino, and 162 remained as "Other APA."

Statistical methods

The basic approach is given in Gail et al. (3). First we developed a multivariate relative risk model from the Asian American Breast Cancer Study data applied to the risk factors in Gail et al. (3). Then we obtained baseline age-specific breast cancer incidence rates by multiplying age- and ethnicity-specific rates from SEER times one minus the common population attributable risk estimated from the Asian American Breast Cancer Study. Finally we made absolute risk projections for an APA woman with specific risk factors by multiplying her multivariate relative risk times the baseline age- and ethnicity-specific breast cancer incidence rate and taking age- and ethnicity-specific competing risks into account. Further details follow.

Age at diagnosis was used for cases. A comparable age was assigned to controls as follows. The mean difference between the date of interview and the date of diagnosis was computed for cases within strata defined by ethnicity, study location, year of birth in 5-year intervals, and age at interview category (above and below the median age of cases at interview). This

mean difference was subtracted from the age at interview of each control woman in that stratum to obtain a comparable age for each control.

Initially, ethnicity-specific odds ratios were obtained using logistic regression separately for Chinese, Japanese, and Filipino women in the Asian American Breast Cancer Study with the same independent variables as in (3) (see Table 1), but with age also included as a continuous variable and with dummy variables for location. The log relative odds model included main effects in four variables: age at birth of first live child (AGEFLB), coded as 0, 1, 2, or 3 for ages of younger than 20, 20-24, 25-29 or nulliparous, or older than 29 years, respectively; number of affected first-degree female relatives (NUMREL), coded as 0 or 1 for zero or more than zero based on mother's, sisters', and daughters' histories of breast cancer as of the date of interview; age at menarche (AGEMEN) coded as 0,1, or 2 for age at menarche ≥ 14 , 12-13, or < 12 years; and number of benign surgical and needle breast biopsies (NBIOPS), coded as 0, 1, or 2 for zero, one, or more than one biopsy examinations. To avoid counting the biopsy that led to the diagnosis of breast cancer in a case patient, we excluded biopsies occurring within 3 years of the date of interview, because breast cancer cases could be ascertained and interviewed up to three years after diagnosis. In addition, we excluded any biopsies that occurred at the same age as the breast cancer diagnosis. Unlike previous models (3), there were no interactions between age and NBIOPS or between AGEFLB and NUMREL, and NUMREL had only two levels.

Formal tests of heterogeneity of the log odds ratio parameters for the four risk factors among the Chinese, Japanese, and Filipino women were not statistically significant. We therefore computed common log odds parameters for the covariates in Table 1 by fitting a logistic

regression that included 18 intercepts for the different combinations of ethnicity (3), location (3) and age (<50 years, ≥50 years) as well as age as a continuous variable and the variables in Table 1. The values of the log odds corresponding to variables in Table 1, and their estimated variance/covariance matrix are in Supplemental Table 2.

To compute an attributable risk, AR, that is representative of the entire SEER population of Chinese, Japanese and Filipino women, we defined the weight for Chinese women as

$$w_C = (D_C / d_C)(D_C + D_J + D_F)^{-1}, \quad (1)$$

where D_C is the number of Chinese breast cancer cases in SEER for the years 2000-2005, d_C is the total Chinese breast cancer cases with complete covariate data in the Asian American Breast Cancer Study, and other terms are defined similarly for Japanese (J) and Filipino (F) groups. Weights for Japanese and Filipino women are also defined similarly. The factor $F(t)=1-AR(t)$ for the combined group of age t is given by a weighted version of the formula by Bruzzi et al (11):

$$1 - AR(t) = w_C \sum_{Chinese} \frac{1}{rr} + w_J \sum_{Japanese} \frac{1}{rr} + w_F \sum_{Filipino} \frac{1}{rr}, \quad (2)$$

where the sums of reciprocal estimated relative risks are over the cases of age t with complete data in the various subgroups of the Asian American Breast Cancer Study. This formula was applied separately for cases under age 50 years and for cases 50 years and older. The weights in equations (1) and (2) are proportional to the weights in the Appendix and yield the same results, because the proportionality factor cancels from ratios in the Appendix. Equation (2) also equals the SEER-weighted average of ethnicity-specific estimates of 1 minus attributable risk, $[D_C\{1 - AR_C(t)\} + D_J\{1 - AR_J(t)\} + D_F\{1 - AR_F(t)\}] / (D_C + D_J + D_F)$.

To compute absolute risks, we used the age- and ethnicity-specific invasive breast cancer incidence rates $h^*(t)$ from Supplemental Table 1 and estimated the baseline hazard as $h_1(t)=h^*(t)F(t)$. The hazard $h_2(t)$ of risks of age- and ethnicity-specific mortality from non-breast cancer causes was obtained from Supplemental Table 1. Using formula (6) in Gail et al. (3) with one year interval widths, we combined information on h_1 , h_2 and the relative risk rr to project individualized absolute risk for various initial ages, final ages, and combinations of risk factors.

For a combination of risk factors leading to a relative risk rr compared to a woman with all risk factors at their lowest risk level, we computed the variance of the estimate $rrF(t)$, and confidence intervals on it, from the influence function approach of Graubard and Fears (12) (see Appendix). Regarding h^* and h_2 as known quantities, we estimated the variance of the estimated absolute risk by Taylor series expansion in $rrF(t)$. A logit transformation of the absolute risk was used to obtain symmetric 95% confidence intervals by adding and subtracting 1.96 times the estimated standard error of the logit transform. Finally, the inverse logit transform was applied to these symmetric confidence limits to obtain 95% confidence intervals on the absolute risk. A computer program in SAS (13) is available to compute such confidence limits for any combination of initial and final ages and risk factors.

We prepared a graph that gives approximate confidence intervals by generating confidence limits for a wide range of absolute risks corresponding to various choices of risk factors and risk projection intervals for Chinese, Japanese, Filipino, Hawaiians, Other Pacific Islanders and Other Asian women. We regressed the upper confidence limits calculated from the variance estimates (Appendix) on the absolute risk, $\phi(x)$, and on $\phi^2(x)$. The points to which

the regressions were fitted were chosen to cover a broad range of absolute risks. For each of the 14 starting ages 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, we considered projection intervals of length 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 and 70 years, subject to the constraint that the starting age plus the duration of the projection interval was at most 90 years. This yielded 105 possible age intervals over which projections were to be made. For each such age interval, we computed the absolute risk for each of the 72 possible risk factor combinations, resulting in $105 \times 72 = 7,560$ pairs for each ethnic group. Thus there were $6 \times 7,560 = 45,360$ estimates of absolute risk and corresponding upper and lower confidence limits. The regressions explained 99.1 percent of the variation in upper confidence limits and 98.4 percent of the variation in lower confidence limits. Thus, the loci in Figure 1 each provide a good fit to the calculated confidence limits in these 45,360 scenarios. The coefficients a , b and c in the regressions $a + b\phi(x) + c\phi^2(x)$ were $(-0.0053, 1.6270, -0.4808)$ for the upper confidence limit and $(0.0026, 0.6219, 0.0038)$ for the lower confidence limit.

To assess the calibration of the AABCS model, we checked it in independent data from APA women in the Women's Health Initiative. We performed separate validation studies to test model calibration for Chinese, Japanese, Filipino, Other Asians (excluding the previous three groups), native Hawaiians, and Pacific Islanders (excluding native Hawaiians). For women in various categories, such as Japanese women aged 50-59 years, we computed the probability of developing invasive breast cancer from the AABCS model based on her age at entry, risk factors, and the age she would attain if she survived to the end of the original Women's Health Initiative follow-up on August 15, 2008. The sum of all such probabilities over women in category i was the expected count, E_i , which we compared with the corresponding

observed number of women with incident invasive breast cancer, O_i . In each category, we computed an O/E ratio and a 95% confidence interval (CI) with lower limit $(O/E)\exp(-1.96O^{-1/2})$ and upper limit $(O/E)\exp(+1.96O^{-1/2})$. In addition, p-values for the goodness-of-fit test were calculated within groupings of categories of the breast cancer risk factors including age at entry, age at menarche, number of biopsies, age at first live birth, and number of affected first-degree relatives. The p-values for the goodness-of-fit tests within these groupings were obtained from the chi-square statistic $\Sigma(O-E)^2/E$ with degrees of freedom equal to the number of mutually exclusive and exhaustive categories within the grouping. For a single category, i , the value $(O_i-E_i)^2/E_i$ was compared to a chi-square distribution with one degree of freedom. To summarize results over ethnic subgroups, we added the E and O values for a given exposure category, such as age group 50-59 or number of biopsies, over the six ethnic subgroups.

The concordance statistic or area under the receiver operating curve (AUC) is the probability that a randomly selected case would have a higher projected absolute invasive breast cancer risk than a randomly selected control (14). In order to estimate how much the factors in the AABCS model contributed to discriminatory accuracy for women of a given age, we estimated age-specific concordance statistics in two age intervals (50-59, ≥ 60 years) with data from Women's Health Initiative data and computed the unweighted average of these age-specific concordance estimates. We used the non-parametric estimator in Wieand et al. (15), which accounts for ties and provides estimates of standard errors.

Results

Relative and Attributable Risks

Relative risks and 95% confidence intervals (CI) estimated from the logistic model for APA women in the Asian American Breast Cancer Study are shown in Table 1, which also indicates the number of cases and controls in various risk factor categories in the Asian American Breast Cancer Study, and the corresponding relative risks from BCRAT (1, 3). To obtain multivariate relative risks from Table 1, one multiplies the separate relative risks for AGEMEN, for the combined age and NBIOPS category, and for the combined AGEFLB and NUMREL category. Adjustments for atypical hyperplasia are described in a footnote to Table 1.

The relative risks (Table 1) and log relative risks (Supplemental Table 2) in the AABCS model are similar to those in BCRAT, which is also known as Gail model 2 (1) for AGEMEN and somewhat larger for NBIOPS in women aged 50 years and over. The combined relative risks from AGEFLB and NUMREL were smaller in the AABCS model for some combinations and larger for others. For example, a woman with first birth at age less than 20 years and two affected relatives had larger relative risks from BCRAT, whereas a woman with first birth above age 29 and one affected first degree relative had larger relative risks with AABCS. The conversion factors were $F(t) = 0.4752$ (95% CI: 0.3255-0.6249) for $t < 50$ and $F(t) = 0.5032$ (95% CI: 0.3630-0.6434) for $t \geq 50$ years, which are lower than the corresponding values in NCI's BCRAT, 0.5788 and 0.5788, and reflect higher attributable risks in the AABCS model.

Individualized Absolute Risk Projections for APA Women

Table 2 gives absolute risks for various initial and final ages and various initial relative risks for Chinese American women. This table is repeated (Supplemental Table 3a), together with

similar tables for Japanese, Filipino, native Hawaiians, Pacific Islanders (excluding native Hawaiians) and Other Asian women in Supplemental Tables 3b-3f respectively.

The use of Tables 1 and 2 to make risk projections is best illustrated by example. Suppose one wishes to project invasive breast cancer risk over 30 years for a 30 year old nulliparous Chinese American woman (AGEFLB=2) who began menstruating at age 14 (AGEMEN=0), whose mother but not sister or daughter had breast cancer (NUMREL= 1), and who has had one breast biopsy (NBIOPS =1). It is unknown whether atypical hyperplasia was present. We obtain the woman's initial relative risk by multiplying relative risks corresponding to the factors in Table 1, namely 1.000 (for AGEMEN=0) $\times 1.738$ (for NBIOPS=1) $\times 3.837$ (for AGEFLB=2 and NUMREL=1) = 6.67. As in Gail et al. (3) , we would recommend multiplying by 1.82 if it were known that any biopsy had atypical hyperplasia and by 0.93 if it were known that atypical hyperplasia was absent. The thirty year risk would be 7.52% if the relative risk were 5.0 (Table 2). An approximation can be obtained by linear interpolation as follows: $7.52 + (14.47-7.52)(6.67-5.00)/(10-5) = 9.84\%$. This result is close to the exact calculation of 9.90%. The second term, which adds 2.32%, corrects for the relative risk of 6.67, instead of 5.00.

Confidence Intervals on Risk Projections

A SAS (13) program provides confidence intervals that take into account random variation in estimates of relative and attributable risks from the Asian American Breast Cancer Study data (Appendix). Approximate 95% confidence intervals can be obtained from Figure 1, which shows loci for upper and lower confidence limits, each plotted against the absolute risk projection. The width of the confidence interval increased with increasing absolute risk. The 95% confidence interval computed by the SAS program for the 30 year projection in the

previous example was 6.30% to 15.22%. The regressions in Figure 1 yielded the approximate 95% confidence interval, 6.36% to 15.27%, in good agreement. For most purposes, Figure 1 yields an adequately accurate confidence interval.

Comparisons with NCI Breast Cancer Risk Assessment Tool

To compare risk projections from the AABCS model with those from BCRAT, we plotted 5-year absolute risks from the AABCS model (ordinate) against those from BCRAT for each of the $3 \times 3 \times 12 = 108$ possible relative risks in the BCRAT separately for Chinese women aged 35 (Figure 2a), 50 (Figure 2b) and 70 years (Figure 2c). These figures are repeated for Chinese women in Supplemental Figure 1a, 1b, and 1c, and for Japanese, Filipino, native Hawaiians, and Pacific Islanders (excluding native Hawaiians) and Other Asian women in Supplemental Figures 2(a,b,c), 3(a,b,c), 4(a,b,c), 5(a,b,c) and 6(a,b,c) respectively.

For Chinese women aged 35 years (Figure 2a), estimates from BCRAT exceeded AABCS model estimates in 99 (92%) of 108 risk factor combinations, as indicated by points below the equiangular (45 degree) line. Because women aged 35 usually have small 5-year risks, the differences in absolute risk estimated from the two models were small. For women aged 50 years, BCRAT estimates exceeded AABCS model estimates in 77 (71%) of risk factor combinations (Figure 2b), and for women aged 70 years, BCRAT estimates exceeded the AABCS model estimates in 103 (95%) of risk factor combinations (Figure 2c). Thus BCRAT yielded higher estimates than the AABCS model for most risk patterns in Chinese women. The proportion of risk factor patterns in which BCRAT gave larger projections than the AABCS model depended on age and ethnicity (Table 3 and Supplemental Figures 1-6). For example, BCRAT produced higher projections than the AABCS model in only 48 (44%) of 70 year old Japanese women (Table 3, Supplemental Figure 2). Thus, for some

combinations of risk factors, ages, and ethnicities, the AABCS model projections exceeded those of BCRAT.

Validation with Data from the Women's Health Initiative

The calibration of the AABCS model was assessed using data from 4,031 APA women who entered the Women's Health Initiative without a prior history of breast cancer (5). The average time of follow-up of this cohort was 9.1 years. From the breast cancer risk factor profiles collected at entry, we used the AABCS model to estimate the number of invasive breast cancer cases that would be expected to occur among the Women's Health Initiative APA cohort members. The results of this assessment are presented in Table 4.

Overall, the AABCS model predicted 120.3 cancer cases, but 141 were observed (Table 4). This yielded an observed to predicted ratio of $O/E=1.17$ (95% CI= 0.99 to 1.38) with $p=0.059$ for testing $O/E=1.0$. The model statistically significantly underestimated risk in women who had taken estrogen and progesterone ($p=0.0053$), in women with no family history of breast cancer in first degree relatives ($p=0.0042$), in women in the lowest predicted quintile of risk ($p=0.0020$), and in "Other Asian" women (0.0009). There was an indication of underestimation of risk for Chinese and Filipino women which was not statistically significant. Thus the AABCS model tended to underestimate risk moderately in the Women's Health Initiative population.

Estimates of the age-specific concordance statistic from the Women's Health Initiative data were 0.636(95%CI: 0.554 to 0.718) for women aged 50 to 59 years and 0.592(95%CI: 0.529 to 0.655) for women aged 60 years and older. Thus the average age-specific concordance was 0.614 (95% CI: 0.587 to 0.640).

To compare rates of breast cancer incidence in APA women in the Women's Health Initiative with those expected in the SEER population, we computed the standardized incidence ratio (SIR) from the SEER rates in Supplemental Table 1. Overall, we found SIR = 1.18 (95%CI: 0.98 to 1.42) among women reporting a single ethnic identity and 1.17 (95%CI: 0.98 to 1.39) among women reporting one or more than one ethnic identities. This SIR range of 1.17 to 1.18 probably explains why the O/E ratio of 1.17 was found for the AABCS model, which was calibrated to these SEER rates.

Discussion

We initially used the same breast cancer risk factors and coding as in the original model of Gail et al. (3) to estimate relative risks and attributable risks for APA women in the Asian American Breast Cancer Study, but the final AABCS model was more parsimonious. In particular, interactions between age at first live birth and number of affected first-degree relatives and between age and number of biopsies were omitted, and number of affected first degree relatives was dichotomized (0 versus 1 or more). This model fit the Asian American Breast Cancer Study data well and yielded absolute risk estimates with smaller variance than models with the original coding. By combining relative and attributable risks from the Asian American Breast Cancer Study case-control data with SEER data on ethnicity- and age-specific breast cancer incidence rates and with data from the National Center of Health Statistics on the ethnicity- and age-specific rates of mortality from non-breast cancer causes, we were able to construct the AABCS model to project individualized absolute invasive breast cancer risk for APA women. Using Table 1 and Supplemental Tables 3a-3f, one can estimate such risks over various time intervals for APA women with specified ethnicity, risk factors and age at counseling. Approximate confidence intervals can be obtained from

Figure 1. A SAS (13) program is available to estimate risks and provide 95% confidence intervals. This program can be downloaded from the web site for the Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, <http://dceg.cancer.gov/bb>.

Except for changes anticipated from recoding NUMREL, the relative risk estimates for the AABCS model resemble those from BCRAT (Table 1). This may explain why the average age-specific concordance statistic for the AABCS model, 0.614, was similar to that reported for the original Gail model (16), 0.596.

Our validation study with independent Women's Health Initiative data indicated that the AABCS model tended to underestimate risk in the Women's Health Initiative by about 17% overall, and more so in Chinese, Filipino, and "Other Asian" populations (Table 4).

However, the Women's Health Initiative breast cancer rates were about 18% higher than predicted from SEER rates, with a standardized incidence ratio of 1.18 (95%CI: 0.98 to 1.42) among women reporting a single ethnic identity. Perhaps the Women's Health Initiative rates are higher than expected from SEER rates because participants in the Women's Health Initiative were self-selected to have higher than average risk or because screening for breast cancer was more intense in the Women's Health Initiative than in the general population. Because the AABCS model was calibrated to SEER rates and meant to apply to women in the general population, we do not regard underestimation of breast cancer incidence in the Women's Health Initiative overall as a reason to recalibrate the AABCS model. However, certain features of the validation study indicate a need for further efforts to assess the model and consider recalibration, including the fact that the AABCS model significantly

underestimates risk in the lowest quintile of predicted risk (Table 4). Two thirds of the women in the Asian American Breast Cancer Study were under age 50 years (Table 1), whereas the WHI cohort included post-menopausal women exclusively. Possible differences in the distributions of risk factors by age and differences in the effect sizes of risk factors in pre-menopausal and post-menopausal women may explain some of the differences between AABCS model predictions and observations in WHI.

As described previously (1), BCRAT uses data on white women from the Breast Cancer Detection Demonstration Project to estimate relative risks, data on risk factor distributions from a population-based study of white women to estimate attributable risk, SEER breast cancer incidence rates for white women, and national non-breast cancer mortality rates for white women. Although BCRAT uses race-specific data for African American women (2), and ethnicity-specific SEER rates for Hispanic women, BCRAT uses only data for white women in projecting rates for APA women, and warns the user that estimates are “uncertain”. In contrast, the AABCS model uses data from Chinese, Japanese and Filipino women to estimate relative and attributable risks, and SEER rates specific for Chinese, Japanese, Filipino, native Hawaiian, Other Pacific Islanders, and Other Asian American women. Because the choice of SEER rates has an important impact on risk projections, and because rates in white women exceed those in most Asian American populations, it is not surprising that BCRAT projections tend to exceed AABCS projections in Chinese, Filipino, Other Pacific Islander and Other Asian populations, but not in native Hawaiians and not in all age groups (Table 3, Supplemental Figures 1-6).

The AABCS model has only modest discriminatory accuracy, in line with that of other breast cancer risk prediction models. There is a need to increase discriminatory accuracy by adding strong risk factors, such as the percent areal mammographic density (16). Apart from the need to develop and validate such a model for APA women, the use of such a model would require more expense and effort than obtaining the data on the risk factors in Table 1.

One must be aware of additional limitations of the AABCS model. Confidence intervals are wider for women with large projected risk than for women with small projected risk (Figure 1). The age range of participants in the Asian American Breast Cancer Study was 20 to 55 years. Thus, projections of risk from the AABCS model rely on the assumption that estimated relative and attributable risks from this comparatively young population also apply to older women. The AABCS model, like BCRAT, should be used with caution or avoided for certain special populations. The AABCS model would usually underestimate risk in APA women with a previous history of invasive breast cancer, ductal carcinoma *in situ*, or lobular carcinoma *in situ*, and in women known to be carrying breast cancer causing mutations, such as mutations in the BRCA1 or BRCA2 genes. Likewise, APA women who received substantial doses of radiation to the breast at a young age, as from radiation treatment of Hodgkin lymphoma, are also likely to be at much higher risk than predicted by the AABCS model (17). Based on the Women's Health Initiative validation data, one should be aware that the AABCS model may underestimate risk in APA women with five-year predicted risk under 0.8% (Table 4). Further validation efforts are needed to assess this issue.

Despite these limitations, the AABCS model, unlike BCRAT, is based on ethnicity-specific data for APA women and usually gives smaller estimates of invasive breast cancer risk for

APA women than the currently available Breast Cancer Risk Assessment Tool. While aware of the need for additional validation studies, we recommend the AABCS model for counseling APA women and for designing and determining eligibility for breast cancer prevention trials.

Funding: This work was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute and by Public Health Service grants (U10-CA069651 and U10-CA069974) to the National Surgical Breast and Bowel Project. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U. S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-, 19, 32122, 42107-26, 42129-32, and 44221).

Notes: We thank Mary Pettinger for preparing the WHI data and analyzing standardized incidence ratios for WHI data and Jennifer Boyd-Morin for checking coding and preparing the data analysis files for the Asian American Breast Cancer Study.

Affiliations of authors: Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD (RKM, RGZ,HL, MHG); Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD (RKM); Graduate School of Public Health and Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA (JPC); Women's Health Initiative Clinical Coordinating Center (GLA) and Division of Public Health Sciences (KEM), Fred Hutchinson Cancer Research Center Seattle, WA; Information Management Services, Rockville, MD (DP)

REFERENCES

1. Costantino JP, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J, Wieand HS. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *Journal of the National Cancer Institute* 1999;91(18):1541-1548.
2. Gail MH, Costantino JP, Pee D, Bondy M, Newman L, Selvan M, Anderson GL, Malone KE, Marchbanks PA, McCaskill-Stevens W, Norman SA, Simon MS, Spirtas R, Ursin G, Bernstein L. Projecting individualized absolute invasive breast cancer risk in African American women. *Journal of the National Cancer Institute* 2007;99:1782-1792.
3. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;81(24):1879-86.
4. Ziegler RG, Hoover RN, Pike MC, Hildesheim A, Nomura AMY, West DW, WuWilliams AH, Kolonel LN, Hornross PL, Rosenthal JF, Hyer MB. Migration patterns and breast-cancer risk in Asian-American women. *Journal of the National Cancer Institute* 1993;85(22):1819-1827.
5. Chlebowski RT, Chen Z, Anderson GL, Rohan T, Aragaki A, Lane D, Dolan NC, Paskett ED, McTiernan A, Hubbell FA, Adams-Campbell LL, Prentice R. Ethnicity and breast cancer: Factors influencing differences in incidence and outcome. *Journal of the National Cancer Institute* 2005;97(6):439-448.
6. Miller BA, Chu KC, Hankey BF, Ries LAG. Cancer incidence and mortality patterns among specific Asian and Pacific Islander populations in the US. *Cancer Causes and Control* 2008;19(3):227-256.
7. Johnson CH, editor. SEER Program Coding and Staging Manual, 2004. Bethesda: National Cancer Institute; 2004.
8. North American Association of Central Cancer Registries Race and Ethnicity Work Group. Asian Pacific Islander Identification Algorithm [NAPIIA v1.2]. Springfield, IL; 2009.
9. Falkenstein MR. The Asian and Pacific Islander surname list: as developed from Census 2000. In: *Proceedings of the Survey Research Methods Section: American Statistical Association*; 2002. p. 946-950.
10. Lauderdale DS, Kestenbaum, B. Asian American ethnic identification by surname. *Population Research and Policy Review* 2000;19:283-300.
11. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk-factors using case-control data. *American Journal of Epidemiology* 1985;122(5):904-913.
12. Graubard BI, Fears TR. Standard errors for attributable risk for simple and complex sample designs. *Biometrics* 2005;61(3):847-855.
13. SAS. SAS 9.2 Software. Cary, North Carolina: SAS Institute Inc.; 2010.
14. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *Journal of the National Cancer Institute* 2001;93(5):358-366.
15. 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.
16. Chen JB, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, Benichou J, Gail MH. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *Journal of the National Cancer Institute* 2006;98(17):1215-1226.

17. Travis LB, Hill D, Dores GM, Gospodarowicz M, van Leeuwen FE, Holowaty E, Glimelius B, Andersson M, Pukkala E, Lynch CF, Pee D, Smith SA, Van't Veer MB, Joensuu T, Storm H, Stovall M, Boice JD, Gilbert E, Gail MH. Cumulative absolute breast cancer risk for young women treated for Hodgkin lymphoma. *Journal of the National Cancer Institute* 2005;97(19):1428-1437.

LEGENDS

Figure 1. Upper and lower 95% confidence limits on the absolute risk plotted against absolute risk.

Figure 2: Five-year absolute risk projections from the Asian American Breast Cancer Study (AABCS) model and the NCI Breast Cancer Risk Assessment Tool (BCRAT) for Chinese women aged 35 years (2a), 50 years (2b) and 70 years (2c).

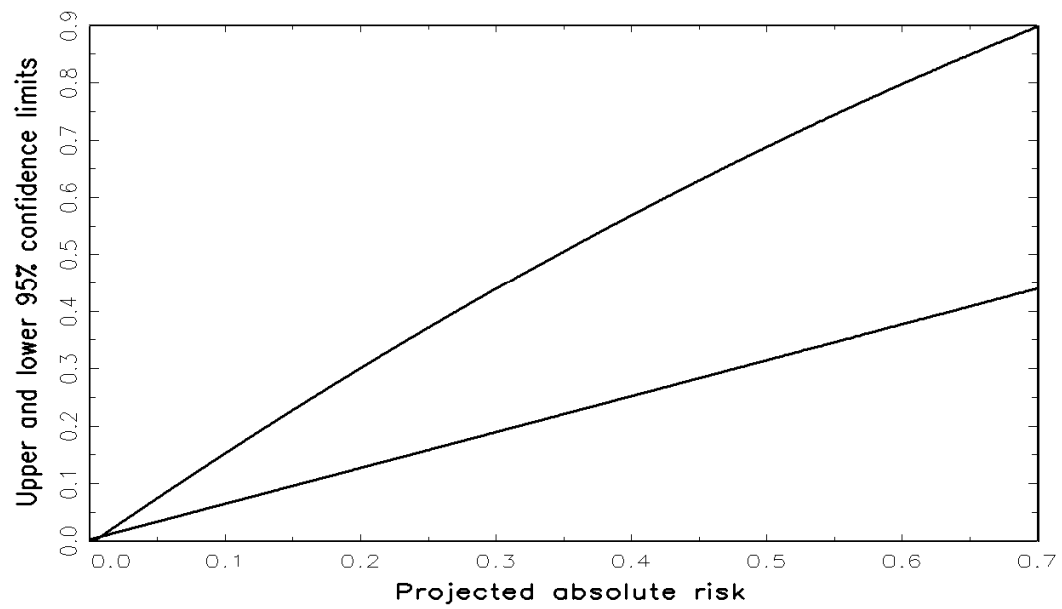


Figure 1. Upper and lower 95% confidence limits on the absolute risk plotted against absolute risk.

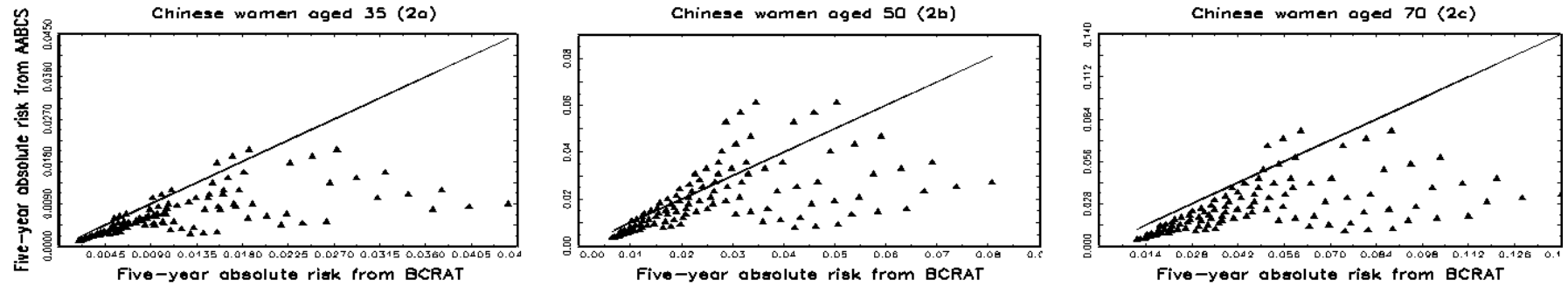


Figure 2: Five-year absolute risk projections from the Asian American Breast Cancer Study (AABCS) model and the NCI Breast Cancer Risk Assessment Tool (BCRAT) for Chinese women aged 35 years (2a), 50 years (2b) and 70 years (2c).

Appendix: Method Used to Calculate Confidence Limits on Absolute Risk Estimates

Let a be the age at the beginning of the risk projection interval and τ be the duration of the risk projection interval. The absolute risk from ages a to $a + \tau$ of an Asian-American woman with risk factors X^* and ethnicity E (1 for Chinese, 2 for Japanese and 3 for Filipino) is given by $\pi_E = \int_a^{a+\tau} h_{1,E}^*(t) \{I_1(t)H_1 + I_2(t)H_2\} \exp(-\int_a^t [h_{1,E}^*(u) \{I_1(u)H_1 + I_2(u)H_2\} + h_{2,E}(u)] du) dt$ where $I_1(t) = 1$ for ages $t < 50$ and $I_1(t) = 0$ otherwise; $I_2(t) = 1$ for $t \geq 50$ and $I_2(t) = 0$ otherwise; $H_1 = \hat{F}_1 \exp(\hat{\beta}X^*)$ where \hat{F}_1 is an estimate of 1-attributable risk for $t < 50$; $H_2 = \hat{F}_2 \exp(\hat{\beta}X^*)$ where \hat{F}_2 is an estimate of 1-attributable risk for $t \geq 50$; $\hat{\beta}$ is an estimate of the log RR of the Gail covariates excluding any intercepts. Both estimates are obtained from the Asian-American case-control data set; $h_{1,E}^*(t)$ is the SEER breast cancer incidences for each ethnicity and $h_{2,E}(t)$ is the competing hazard excluding death from breast cancer for ethnicity E .

We assume that $h_{1,E}^*(t)$ and $h_{2,E}(t)$ are known without error. The variance of the absolute risk π_E is obtained from the delta method as, $D'\Phi D$ where $D' = (\partial\pi_E / \partial H_1, \partial\pi_E / \partial H_2)$ and Φ is the covariance of $(H_1, H_2)'$. Confidence intervals on π_E are obtained by putting symmetric confidence intervals on $\ln(\pi_E / (1 - \pi_E))$ and transforming back to limits on π_E .

Before describing estimation of Φ , we first define weights needed for this calculation. We consider 36 strata which are defined by cross-classifications of case-control status Y (0 for control and 1 for case), age group T (1 for age < 50 and 0 for age ≥ 50), ethnicity E (1 for Chinese, 2 for Japanese and 3 for Filipino) and location L (1 for Hawaii, 2 for San Francisco and 3 for Los Angeles). The weight for the j^{th} subject in the stratum with case-control status y , age group t , ethnicity e and location l is denoted by w_{ytelj} . For controls we have $w_{0telj} = 1$ without regard to their age, ethnicity or location. We want the proportions of three ethnicity groups among cases to be the same as the respective proportions in cases in SEER for age groups 1 and 2 separately. Let P_{te} be the number of Asian American women cases in SEER

with ethnic group e and age group t ; then $P_t = \sum_{e=1}^3 P_{te}$ is the total Asian American women cases in SEER with age group t . Then, for cases we have

$$\text{weights } w_{1telj} = \frac{P_{te} / P_t}{n_{1te} / n_{1t..}}, j = 1, 2, \dots, n_{1tel}, \text{ where } n_{1te} = \sum_{l=1}^3 n_{1tel}, n_{1t..} = \sum_{e=1}^3 \sum_{l=1}^3 n_{1tel} \text{ and } n_{ytel}$$

is the number of subjects in this stratum. In particular, n_{1tel} is the number of cases with complete risk factor data. The sum of case weights for age group t equals $n_{1t..}$. In our data, $P_{11} = 997, P_{12} = 546, P_{13} = 1187, n_{111.} = 105, n_{112.} = 137, \text{ and } n_{113.} = 150$ for women under age 50 years. Likewise, for women aged ≥ 50 years, $P_{21} = 1655, P_{22} = 2344, P_{23} = 2423, n_{211.} = 57, n_{212.} = 102, \text{ and } n_{213.} = 38$.

We applied the influence function method given by Graubard and Fears (12) to estimate Φ . For women aged < 50 years, we have $H_1 = S_1 / S_2$, where

$$S_1 = \sum_{e=1}^3 \sum_{l=1}^3 \sum_{j=1}^{n_{11el}} w_{11elj} \exp\{-\hat{\beta}(X_{11elj} - X^*)\} \text{ and}$$

$$S_2 = \sum_{e=1}^3 \sum_{l=1}^3 \sum_{j=1}^{n_{11el}} w_{11elj}.$$

In the formula, X_{11elj} is the vector of covariates for the j^{th} subject in the stratum with the location and ethnicity specific intercept set to zero. X^* is the corresponding covariate (with intercept zero) for a women whose risk is to be projected. By setting $X^* = 0$ in the expression for H_1 , we obtain \hat{F}_1 , an estimate of the common (1-attributable risk). Because $\hat{\beta}$ is based on the data from all cases and controls, every subject makes a contribution to H_1 and to the analogous quantity for women aged ≥ 50 , namely H_2 .

The influence of observation j in the stratum with case-control status y , age group t , ethnicity e and location l on H_1 is

$$Z_{ytelj} = \begin{cases} S_2^{-1} [\Delta_{11elj}(S_2) - H_1 \Delta_{11elj}(S_2)] & y = 1 \text{ and } t = 1 \\ \frac{\partial S_1}{\partial \beta} \Delta_{ytelj}(\beta) / S_2 & \text{otherwise} \end{cases}$$

where

$$\Delta_{11elj}(S_1) = w_{11elj} \exp\{-\hat{\beta}(X_{11elj} - X^*)\} + \frac{\partial S_1}{\partial \beta} \Delta_{11elj}(\beta) \text{ and}$$

$$\Delta_{ytelj}(\beta) = \left\{ \sum_{y,t,e,l,j} X_{ytelj} X'_{ytelj} P_{ytelj} (1 - P_{ytelj}) \right\}^{-1} X_{ytelj} (y - P_{ytelj}) \text{ with } P_{ytelj} = \exp(\hat{\beta} X_{ytelj}) / \{1 + \exp(\hat{\beta} X_{ytelj})\}. \text{ In}$$

this expression for P_{ytelj} , the intercept term is included in X_{ytelj} . Also

$$\frac{\partial S_1}{\partial \beta} = -\sum_{e,l,j} w_{11elj} (X_{11elj} - X^*) \exp\{-\hat{\beta}(X_{11elj} - X^*)\} \text{ and } \Delta_{11elj}(S_2) = w_{11elj}.$$

Similar influences C_{ytelj} can be calculated for H_2 .

The pairs (Z_{ytelj}, C_{ytelj}) for $y = 0$ or 1 , $t = 1$ or 2 , $e = 1, 2$ or 3 , $l = 1, 2$ or 3 , and

$j = 1, 2, \dots, n_{ytel}$ can be used to find the variances of H_1 and H_2 and their covariance by

summing over stratum-specific variance contributions. For example the estimated variance

of H_1 is $\sum_{y=0}^1 \sum_{t=1}^2 \sum_{e=1}^3 \sum_{l=1}^3 \frac{n_{ytel}}{n_{ytel} - 1} \sum_{j=1}^{n_{ytel}} (Z_{ytelj} - \bar{Z}_{ytel})^2$ where \bar{Z}_{ytel} is the stratum mean.

The covariance between H_1 and H_2 is estimated as

$$\sum_{y=0}^1 \sum_{t=1}^2 \sum_{e=1}^3 \sum_{l=1}^3 \frac{n_{ytel}}{n_{ytel} - 1} (Z_{ytelj} - \bar{Z}_{ytel})(C_{ytelj} - \bar{C}_{ytel}).$$

Table 1. Relative risks estimated from the Asian American Breast Cancer Study for all ethnicities combined and relative risks in the NCI's Breast Cancer Risk Assessment Tool (BCRAT or Gail model (3))

<u>Risk factor</u>		<u>Associated</u> <u>Relative risk (95%</u> <u>CI)*</u>	<u>Gail RR</u>	<u>No. of cases</u> <u>(n=589)†</u>	<u>No. of</u> <u>controls</u> <u>(n=952)†</u>
AGEMEN (yr)					
14+ (0)		1.000	1.000	201	337
12-13 (1)		1.078(0.920-1.263)	1.099	283	455
<12 (2)		1.162(0.846-1.596)	1.207	105	160
NBIOPS					
Age < 50 yr					
0 (0)		1.000	1.000	316	578
1 (1)		1.738(1.381-2.186)	1.698	46	51
2+ (2)		3.020(1.908-4.781)	2.882	30	13
Age 50+ yr					
0 (0)		1.000	1.000	166	275
1 (1)		1.738(1.381-2.186)	1.273	22	32
2+ (2)		3.020(1.908-4.781)	1.62	9	3
AGEFLB (yr)	NUMREL				
<20 (0)	0 (0)	1.000	1.000	14	49
	1 (1)	2.207(1.454-3.351)	2.607	1	1
	2+ (2)	2.207(1.454-3.351)	6.798	0	0
20-24 (1)	0 (0)	1.318(1.145-1.518)	1.244	116	264
	1 (1)	2.910(1.876-4.514)	2.681	14	10
	2+ (2)	2.910(1.876-4.514)	5.775	1	0
25 – 29 or nulliparous (2)	0 (0)	1.738(1.310-2.306)	1.548	280	436
	1 (1)	3.837(2.325-6.332)	2.756	29	23
	2+ (2)	3.837(2.325-6.332)	4.907	2	1
30+ (3)	0 (0)	2.291(1.500-3.501)	1.927	120	160
	1 (1)	5.058(2.801-9.135)	2.834	10	8
	2+ (2)	5.058(2.801-9.135)	4.169	2	0

* To obtain the combined relative risk, multiply the Asian American Breast Cancer Study relative risks for AGEMEN, for the appropriate combination of age and NBIOPS, and for the appropriate combination of NUMREL and AGEFLB. If it is known that atypical hyperplasia was present on any biopsy, multiply the result by 1.82. If it is known that there was no atypical hyperplasia on any biopsy and there was at least one biopsy, multiply the result by 0.93. AGEFLB=age at first live birth; AGEMEN=age at menarche; NBIOPS=number of biopsies; NUMREL= number of affected mother or sisters with breast cancer.

†These counts reflect cases and controls with complete risk factor data, which was used to estimate log-odds ratios and attributable risks.

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

Initial age (years)	Years of follow-up	Relative Risk			
		1	2	5	10
20	5	0	0	0	0.01
	10	0.01	0.02	0.06	0.12
	20	0.17	0.35	0.86	1.72
	30	0.73	1.46	3.61	7.09
30	5	0.04	0.09	0.22	0.45
	10	0.16	0.32	0.80	1.60
	20	0.72	1.44	3.56	6.99
	30	1.55	3.08	7.52	14.47
40	5	0.22	0.43	1.08	2.14
	10	0.56	1.12	2.79	5.49
	20	1.4	2.77	6.79	13.11
	30	2.22	4.40	10.63	20.11
50	5	0.36	0.71	1.77	3.5
50	10	0.84	1.68	4.14	8.12
	20	1.68	3.33	8.12	15.57
	30	2.49	4.92	11.84	22.23
60	5	0.42	0.83	2.07	4.09
60	10	0.86	1.71	4.21	8.25
	20	1.69	3.35	8.16	15.63
70	5	0.45	0.89	2.22	4.39
	10	0.88	1.75	4.32	8.46

Table 3. Numbers (and percentages) of risk patterns for which NCI's Breast Cancer risk Assessment Tool (BCRAT) model projects higher five-year breast cancer risk than the Asian American Breast Cancer Study (AABCS) model, by age and ethnicity*

Age(years)	Chinese	Japanese	Filipino	Hawaiian	Other Pacific Islander	Other Asian
35	99(92%)	99(92%)	99(92%)	36(33%)	90(83%)	102(94%)
50	77(71%)	25(23%)	41(38%)	17(16%)	51(47%)	81(75%)
70	103(95%)	48(44%)	83(77%)	21 (19%)	92(85%)	108(100%)

*There are 108 risk patterns.

Table 4. Comparison of observed numbers (O) of invasive breast cancer cases among all Asian and Pacific Islander women in the Women's Health Initiative population with expected cases computed from the Asian American Breast Cancer Study model (E)

Variable and categories	Number of women with follow-up	Number of cases		O/E(95% CI)	p-value*
		O	E		
All women	4,031	141	120.3	1.17(0.99 to 1.38)	0.059
Age at study entry					
50 - 59	1,416	50	42.0	1.19(0.90 to 1.57)	0.132
60 - 69	1,701	66	51.7	1.28(1.00 to 1.63)	
70 -79	914	25	26.7	0.94(0.63 to 1.39)	
Age at menarche					
≥ 14	1,112	36	28.3	1.27(0.92 to 1.76)	0.258
12 - 13	2,085	72	64.4	1.12(0.89 to 1.41)	
≤ 11	834	33	27.7	1.19(0.85 to 1.68)	
Breast biopsies at study entry					
None	3,319	100	82.5	1.21(1.00 to 1.48)	0.008
One	524	34	23.8	1.43(1.02 to 2.00)	
Two or more	188	7	14.1	0.50(0.24 to 1.04)	
Hormone use					
None	1,875	54	52.0	1.04(0.80 to 1.36)	0.047
Estrogen only	983	33	31.3	1.05(0.75 to 1.48)	
Estrogen and Progesterone	1,173	54	37.0	1.46(1.12 to 1.90)	
Hysterectomy at study entry					
No	2,624	95	77.0	1.23(1.01 to 1.51)	0.111
Yes	1,407	46	43.4	1.06(0.79 to 1.42)	
Race/ethnicity					
Chinese	822	22	16.9	1.30(0.86 to 1.98)	0.029
Japanese	2,260	91	85.3	1.07(0.87 to 1.31)	
Filipino	332	10	8.0	1.25(0.67 to 2.32)	
Native Hawaiian	51	2	2.4	0.84(0.21 to 3.35)	
Other Pacific Islander	25	0	0.6	-----	
Other Asian	541	16	7.2	2.24(1.37 to 3.65)	

**5-yr Predicted Breast
Cancer Risk (%) Quintiles**

≤ 0.796	798	18	8.8	2.04(1.28 to 3.24)	
0.797 - 1.157	804	21	15.0	1.47(0.96 to 2.23)	
1.158 - 1.543	813	21	20.0	1.05(0.68 to 1.61)	
1.544 - 2.046	806	32	26.2	1.18(0.83 to 1.68)	
≥ 2.047	810	49	50.2	0.98(0.74 to 1.29)	0.017

Age at first live birth

≤ 19 or unknown	645	14	11.2	1.25(0.74 to 2.11)	
20 – 24	1,144	37	29.4	1.26(0.91 to 1.74)	
25 – 29	1,735	64	57.9	1.10(0.86 to 1.41)	
≥ 30	507	26	21.8	1.19(0.81 to 1.75)	0.391

**First degree relatives with
history of breast cancer at
study entry**

None	3,547	118	90.8	1.30(1.09 to 1.56)	
One or more	484	23	29.6	0.78(0.52 to 1.17)	0.008

*P-value tests goodness-of-fit over all the categories. O=observed cases; E=expected cases; CI=confidence interval.

Supplement

Supplemental Table 1: Invasive Breast Cancer Incidence Rates (per 10⁵) and Non-Breast Cancer Mortality Rates for Chinese, Japanese, Filipino, native Hawaiians, Pacific Islander (excluding native Hawaiians), and Other Asian women*

<u>Incidence</u>					<u>Other Pacific</u>	<u>Other</u>
	<u>Chinese</u>	<u>Japanese</u>	<u>Filipino</u>	<u>Hawaiian</u>	<u>Islander</u>	<u>Asian</u>
Age						
20 - 24	0.4	0.0	0.8	4.5	0.0	1.2
25 - 29	4.6	9.9	8.1	9.9	7.2	6.0
30 - 34	18.8	28.7	22.7	34.0	28.9	18.4
35 - 39	49.3	54.5	55.0	85.3	60.2	45.5
40 - 44	91.4	115.2	112.9	166.9	75.6	79.1
45 - 49	147.2	185.9	181.4	255.3	76.6	104.8
50 - 53	142.1	260.6	222.4	332.2	189.3	137.2
55 - 59	197.1	322.2	268.0	537.3	236.6	149.5
60 - 63	167.5	400.7	289.1	523.8	284.4	164.7
65 - 69	182.2	352.2	253.4	558.2	292.1	147.8
70 - 73	183.4	359.3	245.7	567.7	233.0	121.6
75 - 79	192.0	358.9	228.7	651.3	203.6	106.8
80 - 83	223.3	353.9	181.5	388.9	148.3	137.6
85 - 89	224.7	205.2	175.1	294.9	101.2	66.2
<u>Mortality</u>						
Age						
20 - 24	21.1	17.4	22.9	56.4	46.6	21.3
25 - 29	19.3	29.6	26.3	37.0	60.0	24.2
30 - 34	24.4	22.8	31.5	102.0	85.1	30.2
35 - 39	31.8	36.3	39.4	123.4	147.8	36.9
40 - 44	47.3	59.1	64.8	209.8	193.1	54.3
45 - 49	80.0	108.6	117.0	298.3	386.7	89.4
50 - 53	121.7	186.0	180.9	540.2	492.5	151.5
55 - 59	210.0	321.7	261.4	959.1	817.7	257.5
60 - 63	343.7	471.9	448.3	1631.5	863.8	432.4
65 - 69	609.7	853.5	739.4	2015.2	1897.5	742.0
70 - 73	1066.5	1243.4	1223.3	2735.5	2925.8	1325.2
75 - 79	2014.9	2023.0	2112.7	5044.7	3840.9	2229.1
80 - 83	3799.1	3772.5	3793.7	7226.2	5287.0	4174.7
85 - 89	9833.4	10614.9	8513.9	14584.5	7474.6	8748.6

* From the Detailed Asian/Pacific Islander Database for the years 1998-2002 (2000-centered), as described in reference 6 of the paper.

Supplemental Table 2. Common log-odds estimates for the combined Chinese, Japanese and Filipino women in the Asian American Breast Cancer Study and their covariance estimates from the logistic model*

	NBIOPS	AGEMEN	AGEFLB	NUMREL
Parameter estimates	0.5526	0.0750	0.2764	0.7919
Covariance estimates	1.3868	-0.0236	-0.0256	-0.0123
		0.6625	0.0449	-0.0620
			0.5252	-0.0200
				4.5813

*The covariance estimates are 10^{-2} times the numbers shown. Unweighted logistic regression was fit to the Asian American Breast Cancer Study data by maximum likelihood, with independent variable codes defined in the Methods section. The model also included continuous age and 18 intercepts corresponding to strata defined by ethnicity (Chinese, Japanese, Filipino), location (San Francisco-Oakland, Los Angeles County, and Oahu, Hawaii), and age (<50 years versus \geq years).

Supplemental Table 3a. Projected absolute risk (%) of developing breast cancer within 5, 10, 20 or 30 years, by relative risk, initial age and years of follow-up for Chinese American women.

Initial age (years)	Years of follow-up	Relative Risk			
		1	2	5	10
20	5	0.00	0.00	0.00	0.01
	10	0.01	0.02	0.06	0.12
	20	0.17	0.35	0.86	1.72
	30	0.73	1.46	3.61	7.09
30	5	0.04	0.09	0.22	0.45
	10	0.16	0.32	0.80	1.60
	20	0.72	1.44	3.56	6.99
	30	1.55	3.08	7.52	14.47
40	5	0.22	0.43	1.08	2.14
	10	0.56	1.12	2.79	5.49
	20	1.40	2.77	6.79	13.11
	30	2.22	4.40	10.63	20.11
50	5	0.36	0.71	1.77	3.50
50	10	0.84	1.68	4.14	8.12
	20	1.68	3.33	8.12	15.57
	30	2.49	4.92	11.84	22.23
60	5	0.42	0.83	2.07	4.09
60	10	0.86	1.71	4.21	8.25
	20	1.69	3.35	8.16	15.63
70	5	0.45	0.89	2.22	4.39
	10	0.88	1.75	4.33	8.46

Supplemental Table 3b. Projected absolute risk (%) of developing breast cancer within 5, 10, 20 or 30 years, by relative risk, initial age and years of follow-up for Japanese American women.

Initial age (years)	Years of follow-up	Relative Risk			
		1	2	5	10
20	5	0.00	0.00	0.00	0.00
	10	0.02	0.05	0.12	0.24
	20	0.22	0.44	1.10	2.18
	30	0.93	1.84	4.54	8.87
30	5	0.07	0.14	0.34	0.68
	10	0.20	0.39	0.98	1.96
	20	0.90	1.80	4.44	8.68
	30	2.31	4.57	11.04	20.85
40	5	0.27	0.55	1.36	2.70
	10	0.71	1.42	3.50	6.88
	20	2.13	4.21	10.19	19.33
	30	3.85	7.56	17.81	32.39
50	5	0.65	1.30	3.21	6.32
50	10	1.44	2.86	6.99	13.48
	20	3.19	6.28	14.96	27.64
	30	4.67	9.11	21.18	37.71
60	5	0.99	1.97	4.86	9.48
60	10	1.83	3.62	8.80	16.81
	20	3.36	6.60	15.66	28.77
70	5	0.87	1.74	4.29	8.38
	10	1.67	3.31	8.06	15.44

Supplemental Table 3c. Projected absolute risk (%) of developing breast cancer within 5, 10, 20 or 30 years, by relative risk, initial age and years of follow-up for Filipino American women.

Initial Age (years)	Years of follow-up	Relative Risk			
		1	2	5	10
20	5	0.00	0.00	0.01	0.02
	10	0.02	0.04	0.10	0.21
	20	0.20	0.41	1.02	2.03
	30	0.89	1.78	4.38	8.58
30	5	0.05	0.11	0.27	0.54
	10	0.18	0.37	0.92	1.83
	20	0.87	1.74	4.29	8.40
	30	2.06	4.08	9.89	18.79
40	5	0.27	0.53	1.33	2.64
	10	0.69	1.38	3.42	6.72
	20	1.89	3.74	9.09	17.34
	30	3.15	6.19	14.75	27.29
50	5	0.56	1.11	2.75	5.42
50	10	1.21	2.41	5.92	11.49
	20	2.49	4.92	11.84	22.25
	30	3.48	6.84	16.22	29.73
60	5	0.72	1.43	3.53	6.94
60	10	1.32	2.63	6.44	12.45
	20	2.35	4.65	11.20	21.09
70	5	0.60	1.19	2.95	5.82
	10	1.11	2.20	5.41	10.51

Supplemental Table 3d. Projected absolute risk (%) of developing breast cancer within 5, 10, 20 or 30 years, by relative risk, initial age and years of follow-up for native Hawaiians.

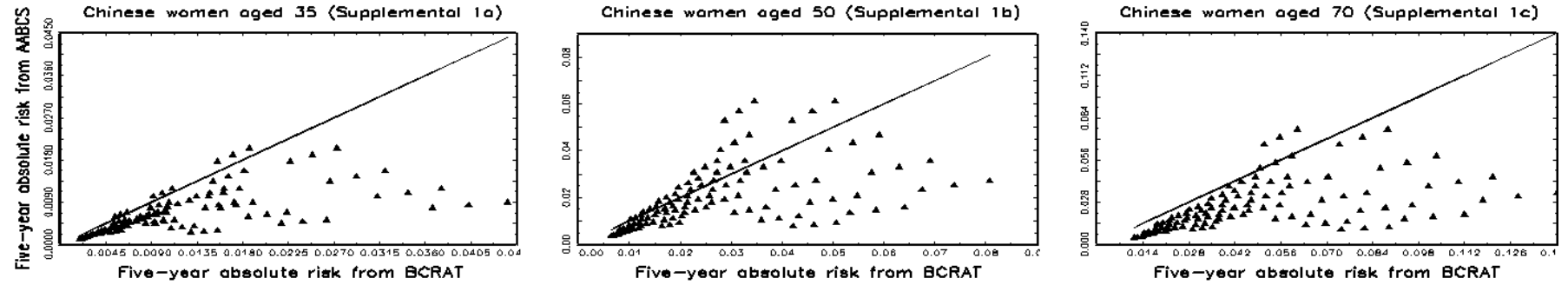
Initial age (years)	Years of follow-up	Relative Risk			
		1	2	5	10
20	5	0.01	0.02	0.05	0.11
	10	0.03	0.07	0.17	0.34
	20	0.31	0.63	1.56	3.09
	30	1.28	2.54	6.24	12.08
30	5	0.08	0.16	0.40	0.80
	10	0.28	0.56	1.40	2.78
	20	1.25	2.49	6.11	11.83
	30	3.24	6.37	15.14	27.91
40	5	0.39	0.79	1.95	3.87
	10	0.99	1.96	4.83	9.42
	20	3.00	5.91	14.10	26.15
	30	5.16	10.03	23.12	40.58
50	5	0.82	1.64	4.04	7.91
50	10	2.09	4.13	10.00	18.97
	20	4.32	8.45	19.72	35.33
	30	6.21	11.99	27.04	46.02
60	5	1.26	2.50	6.13	11.87
60	10	2.46	4.86	11.69	21.95
	20	4.54	8.85	20.50	36.29
70	5	1.33	2.63	6.45	12.46
	10	2.56	5.05	12.10	22.60

Supplemental Table 3e. Projected absolute risk (%) of developing breast cancer within 5, 10, 20 or 30 years, by relative risk, initial age and years of follow-up for Pacific Islander (excluding native Hawaiian) women.

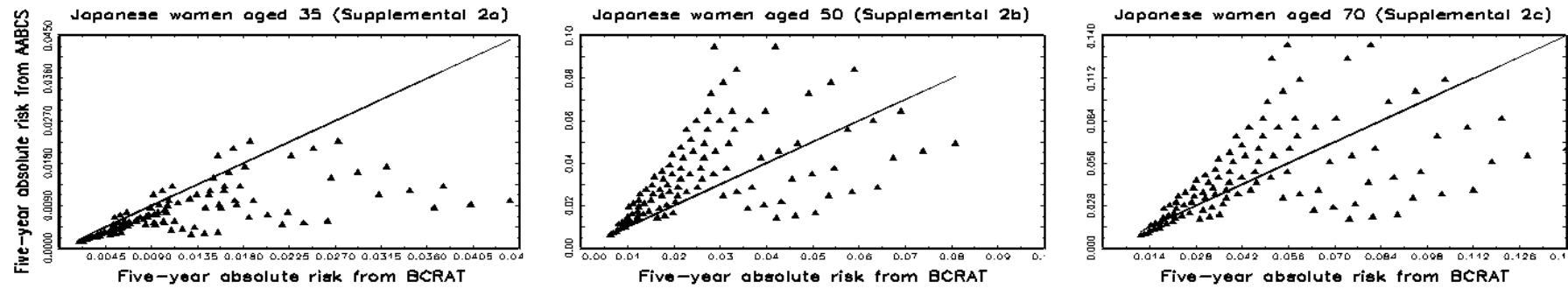
Initial age (years)	Years of follow-up	Relative Risk			
		1	2	5	10
20	5	0.00	0.00	0.00	0.00
	10	0.02	0.03	0.08	0.17
	20	0.23	0.45	1.12	2.24
	30	0.58	1.15	2.85	5.61
30	5	0.07	0.14	0.34	0.68
	10	0.21	0.42	1.05	2.08
	20	0.56	1.12	2.78	5.48
	30	1.55	3.07	7.50	14.42
40	5	0.18	0.36	0.89	1.77
	10	0.36	0.71	1.77	3.51
	20	1.36	2.70	6.60	12.75
	30	2.58	5.09	12.21	22.83
50	5	0.47	0.94	2.32	4.59
50	10	1.03	2.06	5.06	9.86
	20	2.30	4.54	10.94	20.62
	30	3.05	5.99	14.26	26.33
60	5	0.70	1.39	3.44	6.76
60	10	1.36	2.70	6.62	12.79
	20	2.17	4.29	10.36	19.58
70	5	0.54	1.08	2.69	5.30
	10	0.94	1.88	4.62	9.01

Supplemental Table 3f. Projected absolute risk (%) of developing breast cancer within 5, 10, 20 or 30 years, by relative risk, initial age and years of follow-up for Other Asian American women.

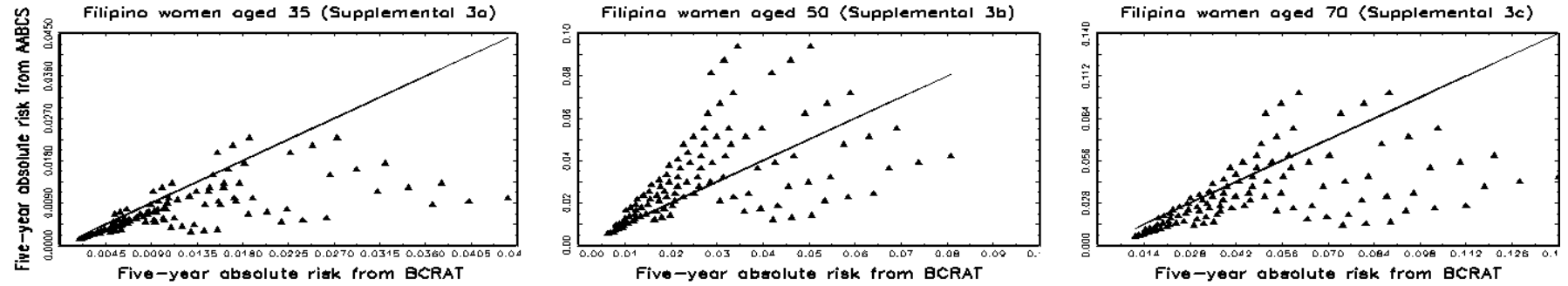
Initial age (years)	Years of follow-up	Relative Risk			
		1	2	5	10
20	5	0.00	0.01	0.01	0.03
	10	0.02	0.03	0.09	0.17
	20	0.17	0.34	0.84	1.67
	30	0.60	1.20	2.96	5.84
30	5	0.04	0.09	0.22	0.44
	10	0.15	0.30	0.75	1.50
	20	0.58	1.16	2.89	5.69
	30	1.28	2.55	6.26	12.12
40	5	0.19	0.37	0.93	1.86
	10	0.43	0.87	2.15	4.26
	20	1.14	2.26	5.57	10.82
	30	1.87	3.71	9.02	17.21
50	5	0.34	0.69	1.71	3.38
50	10	0.71	1.42	3.51	6.90
	20	1.45	2.89	7.07	13.63
	30	1.94	3.84	9.32	17.75
60	5	0.41	0.82	2.03	4.02
60	10	0.76	1.52	3.76	7.38
	20	1.26	2.51	6.15	11.91
70	5	0.30	0.59	1.47	2.92
	10	0.53	1.06	2.63	5.19



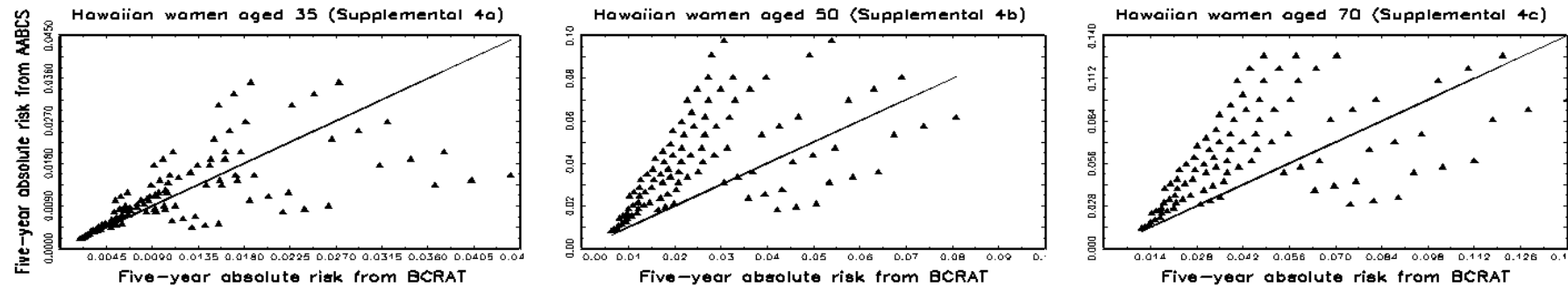
Supplemental Figure 1: Five-year absolute risk projections from the Asian American Breast Cancer Study (AABCS) model and the NCI Breast Cancer Risk Assessment Tool (BCRAT) for Chinese women aged 35 years (1a), 50 years (1b) and 70 years (1c).



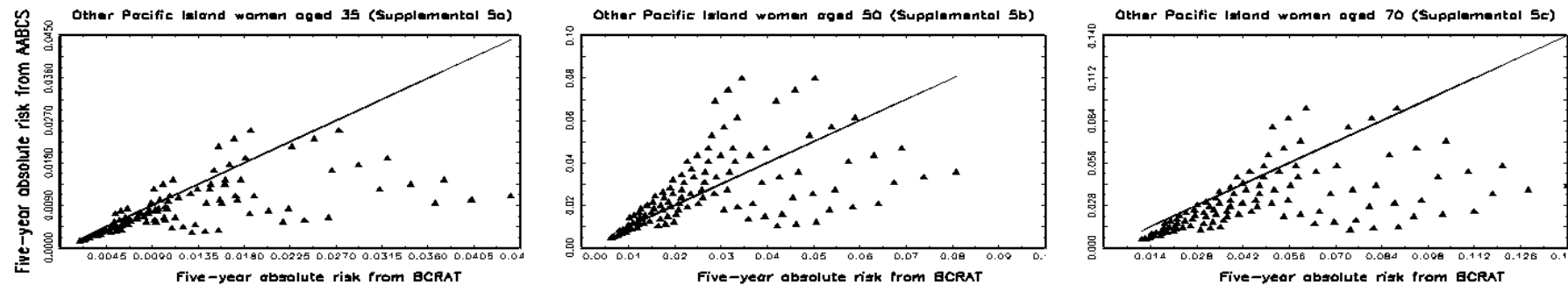
Supplemental Figure 2: Five-year absolute risk projections from the Asian American Breast Cancer Study (AABCS) model and the NCI Breast Cancer Risk Assessment Tool (BCRAT) for Japanese women aged 35 years (2a), 50 years (2b) and 70 years (2c).



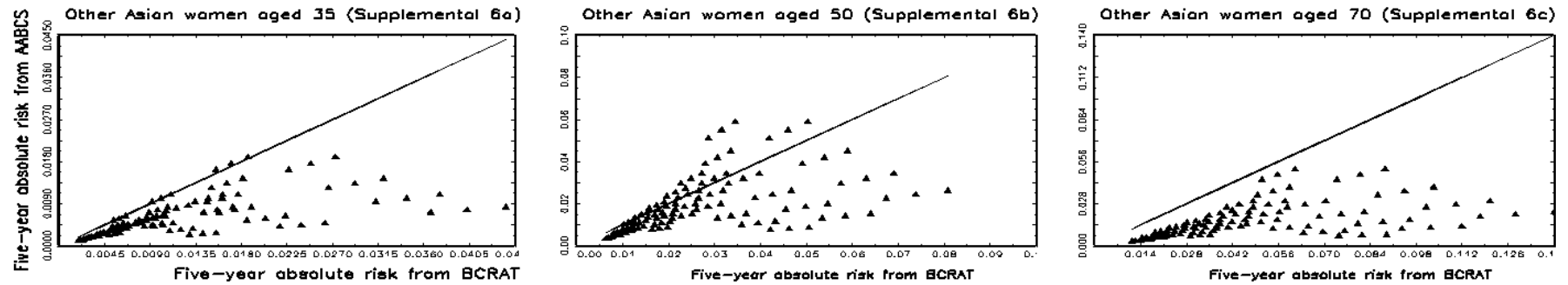
Supplemental Figure 3: Five-year absolute risk projections from the Asian American Breast Cancer Study (AABCS) model and the NCI Breast Cancer Risk Assessment Tool (BCRAT) for Filipino women aged 35 years (3a), 50 years (3b) and 70 years (3c).



Supplemental Figure 4: Five-year absolute risk projections from the Asian American Breast Cancer Study (AABCS) model and the NCI Breast Cancer Risk Assessment Tool (BCRAT) for native Hawaiian women aged 35 years (4a), 50 years (4b) and 70 years (4c).



Supplemental Figure 5: Five-year absolute risk projections from the Asian American Breast Cancer Study (AABCS) model and the NCI Breast Cancer Risk Assessment Tool (BCRAT) for Other Pacific Island women aged 35 years (4a), 50 years (4b) and 70 years (4c).



Supplemental Figure 6: Five-year absolute risk projections from the Asian American Breast Cancer Study (AABCS) model and the NCI Breast Cancer Risk Assessment Tool (BCRAT) for Other Asian American women aged 35 years (4a), 50 years (4b) and 70 years (4c).