Original Study

Assessing Breast Cancer Risk Models in Marin County, a Population With High Rates of Delayed Childbirth

Mark Powell, Farid Jamshidian, Kate Cheyne, Joanne Nititham,
Lee Ann Prebil, Rochelle Ereman

Abstract

Breast cancer risk models have been of limited value for women at high risk. We studied 3 existing models in 12,843 women of the Marin Women's Study where rates of nulliparity and delayed childbirth are very high. The International Breast Intervention Study model performed best, whereas the most widely used Gail model significantly underestimated risk in these women.

Introduction: This study was designed to compare the Breast Cancer Risk Assessment Tool (BCRAT; Gail), International Breast Intervention Study (IBIS; Tyrer-Cuzick), and BRCAPRO breast cancer risk assessment models using data from the Marin Women's Study, a cohort of women within Marin County, California, with high rates of breast cancer, nulliparity, and delayed childbirth. Existing models have not been well-validated in these high-risk populations. **Methods:** Discrimination was assessed using the area under the receiver operating characteristic curve (AUC) and calibration by estimating the ratio of expected-to-observed (E/O) cases. The models were assessed using data from 12,843 participants, of whom 203 had developed cancer during a 5-year period. All tests of statistical significance were 2-sided. **Results:** The IBIS model achieved an AUC of 0.65 (95% confidence interval [CI], 0.61-0.68) compared with 0.62 (95% CI, 0.59-0.66) for BCRAT and 0.60 (95% CI, 0.56-0.63) for BRCAPRO. The corresponding estimated E/O ratios for the models were 1.08 (95% CI, 0.95-1.25), 0.81 (95% CI, 0.71-0.93), and 0.59 (95% CI, 0.52-0.68). In women with age at first birth > 30 years, the AUC for the IBIS, BCRAT, and BRCAPRO models was 0.69 (95% CI, 0.62-0.75), 0.63 (95% CI, 0.56-0.70), and 0.62 (95% CI, 0.56-0.68) and the E/O ratio was 1.15 (95% CI, 0.89-1.47), 0.81 (95% CI, 0.63-1.05), and 0.53 (95% CI, 0.41-0.68), respectively. **Conclusions:** The IBIS model was well calibrated for the high-risk Marin mammography population and demonstrated the best calibration of the 3 models in nulliparous women. The IBIS model also achieved the greatest overall discrimination and displayed superior discrimination for women with age at first birth > 30 years.

Introduction

Within the past 2 decades, increasing interest in breast cancer risk prediction has stimulated the development of several risk models. These models have been used to identify women at high risk of cancer who might benefit from targeted screening or chemoprevention, to estimate the population burden, and to assist physicians and patients in clinical decision-making. However, validation studies of these models have shown variability in calibration and discrimination when applied to differing populations. Many of the models have significantly underestimated the risk in women who are nulliparous or whose first live birth was after the age of 30 years and in other high-risk populations.[1-4] Risk models are best calibrated to the population from which they were developed and are typically developed using general populations to enhance the applicability of the model to outside populations. Thus, the models might have a lower performance in populations with high numbers of either high- or low-risk women.

The County of Marin, just north of San Francisco, has the highest rates of nulliparity and age at first birth > 30 years of all

California counties and has, for many years, had breast cancer rates in the uppermost range of counties in the United States.[5-8] It has been postulated that at least part of the explanation for these increased rates is the delayed childbirth seen in Marin women.[5,8] Of the respondents in the Marin Women's Study (MWS), a mammography-based study of women in Marin County, 57.6% had not had a child by 30 years of age. According to a report by the Centers for Disease Control and Prevention in 2009, the average age at the first birth has increased from 21.4 to 24.0 in the United States from 1970 to 2006, emphasizing the importance of the accuracy of existing models in this population.[9] This trend has not been just limited to the United States and was seen in all developed countries studied, with the United States actually having the lowest age at the of first birth of all developed countries. According to a report from Pew Research, from 1990 to 2008, the percentage of children born to mothers aged $\geq$ 35 years has increased from 9% to 14% and has tripled for women aged $\geq$ 40 years.[10]

Methods

The present study was done as a retrospective cohort within the MWS, and all women without breast cancer as of January 1, 2003 were selected to compare the performance of 3 different risk prediction models during a 5-year follow-up period. The performance of the risk prediction models was assessed using 2 criteria: calibration and discrimination. Calibration is a measure of the ability of a model to accurately predict the number of events in a population. Discrimination measures the model's ability to discriminate at the individual level between women who will and will not develop the event and is measured by calculating the area under the receiver operating characteristic (ROC) curve (AUC). An AUC of 0.5 identifies a model whose discriminatory accuracy is no better than the toss of a coin, and an AUC of 1.0 identifies a model with perfect discriminatory accuracy.

Included in the present analysis were 3 of the most widely used models, the Breast Cancer Risk Assessment Tool (BCRAT or Gail model),[11] BRCAPRO,[12] and International Breast Intervention Study (IBIS) Breast Cancer Risk Evaluation Tool (Tyrer-Cuzick model).[13] These models were tested for calibration and discrimination using the participants of the MWS to assess the performance of these models in a population of women known to have high rates of nulliparity and delayed childbirth.

The Marin General Hospital and Kaiser Permanente Northern California institutional review boards approved the present study, and all participants provided informed consent to fully participate in the study. MWS questionnaire data were collected from 2006 to 2009, the reference baseline was set at the start of 2003, and the population of interest was restricted to women who were breast cancer free at this baseline. The outcome was defined as occurrence of any invasive breast cancer between 2003 and 2007 either reported by women on the questionnaire or included in the cancer registry data obtained from the San Francisco Mammography Registry (SFMR).

Marin Women's Study

Funded by the Centers for Disease Control and Prevention, the MWS was conducted in Marin County, California, at all major screening centers in the county, including those associated with

Kaiser Permanente, Marin General Hospital, and Novato Community Hospitals. These mammography sites are included in the SFMR, 1 of 7 registries participating in the National Cancer Institute Breast Cancer Surveillance Consortium. All women were asked to participate regardless of their history. In the period of enrollment from 2006 to 2009, the MWS enrolled 13,344 women living in Marin County, representing 21.5% of all women of mammography age in the county.[14]

The MWS collected detailed risk factor information, saliva specimens, and the mammographic breast density (both Breast Imaging Reporting and Data System and single x-ray absorptiometry compositional density) from women undergoing mammography at the area facilities. The pathologic findings and case status data were obtained from the SFMR, which collected data from the same women during the same period.

All women enrolled in the MWS were asked to complete an in-depth 87-item questionnaire that included questions regarding reproductive history, the use of exogenous hormones, life course alcohol intake, smoking history, and family history of breast cancer. Additional information collected included current and high school socioeconomic status, diet and nutrition, medications, environmental exposures, measures of stress, education level, work status, and years of residence in Marin County.

Secondary data were obtained by linkage with the SFMR and included compositional breast density (single x-ray absorptiometry), breast cancer case status, demographic data, body mass index (BMI), and family history (including a history of breast cancer in first-degree female relatives and age at diagnosis).

The study population for the present analysis was defined as the subset of women in the Marin County mammography population who were breast cancer free at the beginning of 2003. A total of 501 women were excluded by this criterion, resulting in a subset of 12,843 MWS participants. Of these, 203 women developed invasive breast cancer during the subsequent 5-year follow-up period ending in 2007.

The 3 risk prediction models included in the analysis are described in the subsequent sections, and the variables included in each model are listed in Table 1.

BCRAT (Gail) Model

The Gail model was the first breast cancer risk prediction model to be widely disseminated to both health care professionals for use in clinical settings and the public. The model focuses on nongenetic risk factors, with limited information on family history. The Gail model is unique in that it has been validated in 3 large population-based databases and has been shown to work best in general assessment clinics where family history is not the main reason for referral.[11,15,16] In a recent systematic review by Amir et al,[1] it was reported that although the Gail model was well calibrated, it underestimated the risk in women who were nulliparous or whose first live birth was after age 30 years.[2,17-20]

BRCAPRO Models

The BRCAPRO models[12] were originally developed by Chen et al[21] in 1997 to determine the likelihood of carrying a BRCA gene mutation based on family history. The computerized BRCAPRO model includes an extension software package that enables the