Incidence of Breast Cancer in the United States: Current and Future Trends

William F. Anderson, Hormuzd A. Katki, Philip S. Rosenberg

Background  The incidence of breast cancer increased in the United States until circa 2000 then decreased, mostly among women with estrogen receptor (ER)-positive cancers. Time trends provide important clues for cancer etiology and prevention; however, the observed trends of ER-positive and ER-negative breast cancers can be biased by missing ER data.

Methods  We developed a simple imputation method to correct invasive female breast cancer incidence for missing or unknown ER expression, using nationally representative data from the National Cancer Institute’s Surveillance, Epidemiology, and End Results Program during 1980–2008, including 588,720 invasive female breast cancer patients with 471,336,233 woman-years of follow-up. Corrected rates of ER-positive and ER-negative breast cancers were used to calculate age-standardized incidence rates, estimated annual percentage changes, and projections derived from age–period–cohort models.

Results  The recent decrease in the incidence of breast cancer overall stabilized near 200 per 100,000 woman-years by 2007–2008, reflecting a transient decrease in ER-positive cancers and a steady decrease in ER-negative cancers. The projected incidence rate for breast cancer overall through the year 2016 was similar to the incidence rate during 2007–2008. In contrast, rates of ER-positive breast cancers were projected to increase 5.3% (95% confidence interval = 5.2% to 5.4%), whereas rates of ER-negative breast cancers were projected to decrease 11.4% (95% confidence interval = 11.3% to 11.6%) during 2009–2016.

Conclusion  Recent changes in breast cancer incidence overall reflect the superimposition of divergent trends in ER-positive and ER-negative cancers. If current trends continue, the incidence of ER-positive breast cancers will increase, the incidence of ER-negative breast cancers will continue to decrease, and the incidence of breast cancer overall will remain similar to its current level.

J Natl Cancer Inst 2011;103:1397–1402

Breast cancer is a signature disease of Western populations. It is now known to be composed of at least two main types (1–3) on the basis of a positive or negative test for the presence of the estrogen receptor (ER). The US National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) Program began collecting ER data in 1990 (4,5). Because ER status is both prognostic and predictive (6), the assessment of ER became a universal standard for breast cancer care (7).

Previous SEER reports described dynamic changes in ER-specific breast cancer incidence in the United States from 2000 through 2007 (8–10). One innovation introduced by these studies was the application of statistical methods to account for the possible impact of missing ER data on the estimated trends. However, to date, no one has considered the effects of missing ER data for the entire period of SEER collection, and we hypothesized that the potential impact of the missing data on calculating incidence trends might be substantial.

In this report, we develop a simple statistical method to impute corrected incidence rates of ER-positive and ER-negative breast cancer for missing or unknown ER data. We show striking differences between the apparent and corrected trends between 1992 and 2008. Because of recent large changes in breast cancer incidence, future predictions of rates also have great interest. Therefore, we used corrected rates of ER-positive and ER-negative breast cancer to project incidence from 2009 through 2016.

Materials and Methods

Data  We obtained breast cancer incidence data from the US National Cancer Institute’s SEER Program from January 1, 1980, through December 31, 2008. We used patient and population data from the SEER 9 Registries Database (4) (covering Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-
CONTEXT AND CAVEATS

Prior knowledge

Previous reports using data from the Surveillance, Epidemiology, and End Results Program to determine incidence trends for estrogen receptor (ER)-specific breast cancer in the United States have not considered the impact of missing or unknown ER data.

Study design

Incidence rates of ER-positive and ER-negative breast cancer were calculated using incidence data from the US National Cancer Institute’s Surveillance, Epidemiology and End Results Program from 1992 through 2008 after using a statistical method to account for missing or unknown ER data. Future breast cancer incidence trends were for ER-positive and ER-negative breast cancers were determined on the basis of the corrected incidence rates.

Contribution

Decreases in breast cancer overall observed in the year 2000 did not continue through 2007, but remained steady. Corrected projections indicate that the rate of ER-positive breast cancer will increase from 2009 through 2016, especially for younger women, and that the rate of ER-negative breast cancer will decrease. Similar breast cancer trends were observed among black and non-Hispanic white women.

Implications

The current incidence of breast cancer overall is projected to remain at the current level, but the future rates of hormone-sensitive and hormone-insensitive breast cancers will change in the United States.

Limitations

A statistical model was used to correct data for unknown ER status on the basis of assumptions that ER data have the same chance of being missing among all patients of the same age who are diagnosed in the same year. The divergent trends in ER-positive and ER-negative breast cancer incidences calculated with data corrected for unknown ER status require further studies to reveal the underlying biology.

From the Editors

Puget Sound, and Utah) and the SEER 13 Registries Database (5) (also including Los Angeles, San-Jose Monterey, Rural Georgia, and Alaska Native Tumor Registry) for the registration periods 1980–1991 and 1992–2008, respectively. We assembled a comprehensive dataset of invasive female breast cancers by single years of age at diagnosis \(a\) (aged 30–84 years), calendar year of diagnosis \(t\), ER status (positive, negative, and unknown), and race/ethnicity (non-Hispanic whites, Hispanic whites, blacks, and Asian or Pacific Islanders). ER and race/ethnicity status were recorded in the SEER 13 Registries Database for the contemporary 1992–2008 study period. This study did not include interaction with human subjects or use personal identifying information from the publicly available SEER data, so institutional review board approval and informed consent were not applicable.

Statistical Analysis

Incidence rates were age standardized to the 2000 US standard population by the direct method and expressed per 100,000 woman-years. The overall linear trend in the age-standardized rate was summarized with the estimated annual percentage change (EAPC), computed with Poisson regression for observed rates and a parametric bootstrap for rates that were imputed for missing ER data (see Supplementary Materials, available online). The EAPC is a log-linear model estimator that assumes a constant rate of change. To estimate the linear trend in the age-standardized rate when the constant change assumption might not hold, we supplemented the EAPC with two alternative summary measures, the two-point estimator and the adaptive estimator, previously described by Fay et al. (11).

We developed a simple imputation model to correct the apparent rates of ER-positive and ER-negative breast cancer for missing ER data. For each age \(a\) and year \(t\), we partitioned the observed total number of incident breast cancers \(I^0_a\) according to ER status. Hence

\[
I^0_a = P^0_a + N^0_a + U^0_a
\]  

where \(P^0_a\), \(N^0_a\), and \(U^0_a\) are the observed ER-positive, ER-negative, and unknown counts, respectively. Prior studies have analyzed \(P^0_a\) and \(N^0_a\), but if \(U^0_a\) varies by either \(a\) or \(t\) these results may be biased. Our imputation method estimates the unobserved complete data

\[
I^0_a = P^C_a + N^C_a
\]  

where \(P^C_a\) and \(N^C_a\) are the true numbers of ER-positive and ER-negative counts, respectively.

Our model assumes that unknown ER status is missing at random within a single year of age \(a\) and calendar year \(t\) of diagnosis. Under this model, the observed probability \(\hat{P}_a = P^0_a / (P^0_a + N^0_a)\) on the basis of patients for whom we have complete information is an unbiased estimator of the true probability at the population level that an incident breast cancer diagnosed among women age \(a\) and calendar year \(t\) is ER positive. The equations

\[
\hat{P}^C_a = \hat{P}_a I^0_a
\]

\[
N^C_a = I^0_a - \hat{P}^C_a
\]

provide unbiased estimators of the true numbers of ER-positive and ER-negative breast cancers in the population.

\(P^C_a\) and \(N^C_a\) were used to calculate the age-standardized rate during time periods, the EAPC of the age-standardized rate, and parameters of the age–period–cohort models including the net drift and birth cohort deviations (12–14). Net drift measures the overall log-linear trend by calendar period and birth cohort; cohort deviations quantify departures from the cohort trend that are associated with changes in cancer risk from one generation to the next. A bootstrap procedure was used to assess the 95% confidence intervals (CIs) of these quantities (Supplementary Materials, available online) (15). We applied our method to breast cancer overall (all patients and races combined) and separately for non-Hispanic white, Hispanic white, black, and Asian or Pacific Islander racial or ethnic groups.

With the imputed rates from SEER 13 Registries Database (1992–2008), we projected future breast cancer incidence trends (2009–2016) with age–period–cohort models (16,17). That is, cohort-specific age-at-onset curves (18) and net drifts (12,13) in the age–period–cohort models were used to extend fitted incidence
rates for observed birth cohorts into future calendar periods. Rates for subsequent younger cohorts were extrapolated on the basis of age-at-onset curves with an offset that changed each year by an amount equal to the estimated net drift. Bootstrapped 95% prediction intervals were constructed that incorporated both uncertainty of parameters estimated from the observed data as well as the expected variability of unobserved future cohort and period deviations. Our projections quantify the future implications of observed trends assuming no major future changes in screening or risk factors. All statistical tests were two-sided, and \( P \) values less than .05 were considered statistically significant.

Results

The combined SEER 9 and SEER 13 Registries Databases from 1980 through 2008 included 588 720 invasive female breast cancer patients with 471 362 331 woman-years of follow-up. The SEER 13 Registries Database had 429 757 invasive female breast cancers, including 278 759 ER-positive, 79 865 ER-negative, and 71 133 ER-unknown cancers. Given the relatively low number of breast cancer patients younger than age 30 (3 431), our main analyses focused on women aged 30–84 years.

The proportion of patients with missing ER status statistically significantly declined from 25.9% in 1992 to 5.0% in 2008 (difference = 20.9%, 95% CI = 19.2% to 22.6%, \( P < .001 \)) (Figure 1, A). The proportion of all cancers with known ER status that were reported to be ER-negative was inversely associated with age at diagnosis, decreasing from 44.1% among women diagnosed at ages 30–34 years to 14.9% among women diagnosed at ages 80–84 years (difference = 29.2%, 95% CI = 26.9% to 31.6%, \( P < .001 \)) (Figure 1, B). The corresponding proportion of cancers reported to be ER positive statistically significantly increased from 55.9% for women diagnosed at ages 30–34 years to 85.1% among women diagnosed at ages 80–84 years (difference = 29.2%, 95% CI = 27.4% to 31.1%, \( P < .001 \)).

Seventy-eight percent of breast cancers with missing ER status were imputed to be ER positive and the remainder ER negative. Imputed ER-positive counts were predominant in older age-groups (Supplementary Figure 1, available online), whereas imputed ER-negative counts were predominant in younger age-groups. More imputation was needed for past than recent years; therefore, reassignment of unknown data had greater impact on earlier than recent periods. Consequently, an apparent secular increase in the observed rates of ER-positive breast cancer was attenuated after correction for missing ER data, whereas an apparently stable ER-negative trend decreased after correction (Figure 1, C). Imputed data were used for all subsequent analyses.

We calculated age-standardized incidence rates of breast cancer overall, ER-positive and ER-negative cancers for the past (SEER 9; 1980 through 1991), contemporary (SEER 13; 1992 through 2008) and future (projected from January 1, 2009, through December 31, 2016) periods (Figure 2). The rate of breast cancer overall increased from January 1, 1980, through December 31, 1999, when it peaked at 232 per 100 000 woman-years and then decreased. During the entire contemporary period, a slight downward trend was observed for breast cancer overall. The EAPC was \(-0.39\%\) per year (95% CI = \(-0.78\%\) to 0.01%). The alternative estimators had similar point estimates and overlapping confidence intervals: The two-point estimator was \(-0.15\%\) per year (95% CI = \(-0.26\%\) to \(-0.05\%\)), the adaptive estimator was \(-0.20\%\) per year (95% CI = \(-0.29\%\) to \(-0.05\%\)), and the age-period-cohort net drift was \(-0.24\%\) per year (95% CI = \(-0.32\%\) to \(-0.17\%)\). Given the similarity of all four estimates, we used the more familiar EAPC as the summary measure for all subsequent trends.

Rates of breast cancer overall are projected to stabilize near 200 cancers per 100 000 woman-years from January 1, 2009, through December 31, 2016 (Figure 2), reflecting a projected increase in ER-positive cancers and decrease in ER-negative cancers. Specifically, ER-positive cancers are projected to increase 5.3% (95% CI = 5.2% to 5.4%) during 2009–2016 (from 157.7 to 166.1 per 100 000 woman-years) on the basis of an EAPC of 0.75% per year (95% CI = 0.48% to 1.01% per year). ER-negative cancers are projected to decrease 11.4% (95% CI = 11.3% to 11.6%) during 2009–2016 (from 42.8 to 37.9 per 100 000 woman-years) on the basis of an EAPC of \(-1.69\%\) per year (95% CI = \(-1.71\%\) to \(-1.67\%\) per year).

Among younger women aged 30–49 years (Figure 3, A and Supplementary Table 1, available online), the rate of ER-positive...
Among blacks and non-Hispanic whites, the net drifts and birth cohort deviations in the age-period-cohort models were statistically significantly different ($P < .01$) for both ER-positive and ER-negative cancers (Supplementary Figure 3, available online). Furthermore, these parameters statistically significantly differed by ER status ($P < .001$), demonstrating distinct calendar period effects (eg, screening and/or patient ascertainment) and birth cohort effects (eg, risk factor patterns) for ER-positive vs ER-negative breast cancers within each racial group.

**Discussion**

We developed a simple imputation model to correct the incidence rates of breast cancer for missing ER data. The amount of missing ER data decreased markedly during the period investigated; therefore, more correction was needed at the beginning than at the end of our study period. Consequently, imputation elevated past rates more than recent rates. Imputation was used in previous SEER studies with shorter periods (8–10). Our method is complimentary. The method used in the prior SEER studies makes imputations for individuals, whereas our focus is on corrected counts in aggregate. Nonetheless, our results appear very similar during comparable periods from 2000 through 2007.

Initially, there was hope that the decrease in the incidence of breast cancer overall circa 2000 was a turning point in the long-term increase of breast cancer incidence (19). Unfortunately, our study adds to emerging evidence that this may not be true (10). Indeed, the recent SEER study suggested that decreases in breast cancer overall did not continue through 2007 (10). The analysis of more recent data in our study supports this conclusion and further suggests that the rate of breast cancer overall will remain at the current high level in the near future.

Breast cancer trends are of great interest, but breast cancer overall is a superimposition of ER-positive and ER-negative cancers. A complex pattern exists for ER-positive breast cancer in the United States. It is plausible to attribute the rapid decrease in ER-positive tumors circa 2000 to changes in use of hormone replacement therapy following the Women Health Initiative (WHI) report (8,10,20) and/or to the saturation of screening mammography (21,22). However, the current incidence of imputed ER-positive tumors remains high and is similar to the pre-2000 period. DeSantis et al. (10) recently noted that incidence rates of ER-positive breast cancer had stabilized (overall) or increased (aged 40–49 years) from 2003 through 2007. In fact, our projections suggest that ER-positive cancers will likely increase in the near term (2009 through 2016) and more so for younger than older women.

Conversely, ER-negative rates show a more encouraging trend with a steady decrease of nearly 2% per year. If this current trend continues, we project that ER-negative breast cancers will decrease by an additional 11.4% in the United States from 2009 through 2016. This is certainly good news because ER-negative cancers include the subtypes of breast cancer that are the most difficult to treat (23). Although more sensitive ER tests and/or lower diagnostic thresholds for ER-positive cancers might contribute to the reduction of ER-negative disease (7,10,24), statistically significantly different birth cohort deviations for ER-positive and ER-negative

**Figure 2.** Incidence rates for breast cancer overall and imputed rates of estrogen receptor (ER)–positive and ER-negative breast cancers for early (1980–1991), contemporary (1992–2008), and future (2009–2016) periods. Incidence rates were age standardized to the year 2000 US standard population by the direct method and expressed per 100,000 woman-years among women aged 30–84 years, and 95% confidence intervals are shown (dashed lines). Projected breast cancer rates were obtained using age-period-cohort models.
The primary limitation of our study is that a statistical model is used. Our key model assumption is that ER data have the same chance of being missing among all patients who are diagnosed in the same year at the same age. This working model appears reasonable, given the data in Figure 1, A and B, and that most missing ER reports in SEER reflect administrative omissions rather than ambiguous test results, that is, less than 0.4% of breast cancers were coded as missing because of a test that was not determined to be positive or negative. We also developed an extended imputation model that incorporated age and year of diagnosis as well as the American Joint Committee on Cancer TNM stage (27) and tumor grade (Supplementary Materials, available online). Assignments varied at the individual level, but the overall imputed counts were very similar to our basic model (compare Supplementary Tables 1 and 2, available online). Finally, we obtained very similar estimates of the log-linear trends using four different estimates as follows: 1) EAPC (11), 2) two-point estimator (11), 3) adaptive estimator (11), and 4) net drift (12,13). The similar data produced from the use of these four different estimators give additional support for our current as well as projected trends.

Although the explanations for the divergent trends in ER-positive and ER-negative cancers require further analytical studies, more accurate assessment of the past trends in incidence of breast cancer made on the basis of methods such as ours may help to better gauge the future course of this epidemic malignancy and shed further light on the underlying etiologies and prevention strategies. Decreasing rates of ER-negative breast cancer coupled with increasing rates of ER-positive breast cancer will at least moderate the mixture of hormone-sensitive relative to hormone-insensitive breast cancers. Nonetheless, our projections suggest that in the near future (2009–2016), the incidence of breast cancer overall in the United States will remain close to the currently high level.

References


Funding

This research was supported entirely by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics.

Notes

All of the authors had full access to all of the data in the study and took responsibility for integrity of the data and accuracy of the data analysis. We would like to thank the reviewers for their helpful comments that greatly improved the content of this article.

Affiliation of authors: Division of Cancer Epidemiology and Genetics, Biostatistics Branch, Department of Health and Human Services at the National Cancer Institute, National Institutes of Health, Bethesda, MD.