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

Anderson WF, Katki HA, and Rosenberg PS.

Supplementary Materials

Imputation model

Standard analytically-derived estimators of the variance of age-standardized rates and annual percentage change for the imputed (or corrected) estrogen receptor (ER) data are not readily modified in the presence of missing ER data because these must account for inherent randomness in the observed estimates as well as randomness in the inverse-probability weights

$$\hat{\pi}_{at} = (P_{at}^0 + N_{at}^0) / I_{at}^0 \quad [1]$$

in which P_{at}^0 , N_{at}^0 , and I_{at}^0 are the observed ER positive, ER negative, and total counts, respectively. This randomness will induce correlation between the complete counts for ER positive P_{at}^C and ER negative N_{at}^C breast cancers. Hence, we used a parametric bootstrap to estimate variances and confidence intervals for the imputed data (1).

We assumed that the observed rates for ER positive, ER negative, and ER unknown breast cancers can be described by independent Poisson processes acting on a fixed number of woman-years at risk. We resampled $b = 1,..., B(B = 2000)$ bootstrap replicates of counts for ER positive, ER negative, and ER unknown breast cancers for each age a and calendar year t from independent Poisson distributions

$$P_{at}^b \sim Poisson(P_{at}) , N_{at}^b \sim Poisson(N_{at}) , U_{at}^b \sim Poisson(U_{at}) , \quad [2]$$

in which the mean of each Poisson distribution is the observed count for that age a and calendar year t. Then the total number of cancers in each bootstrap replicate is

$$I_{at}^{b} = P_{at}^{b} + N_{at}^{b} + U_{at}^{b} \quad [3]$$

For each replicate b, we estimated the imputed number of ER positive and ER negative breast cancers (P_{at}^{Cb} and N_{at}^{Cb}, respectively) with inverse-probability weighted estimators (2),

$$\begin{aligned} \hat{P}_{at}^{Cb} &= P_{at}^{b} / \hat{\pi}_{at}^{b} \\ \hat{N}_{at}^{Cb} &= I_{at}^{b} - \hat{P}_{at}^{Cb} \end{aligned} \quad [4]$$

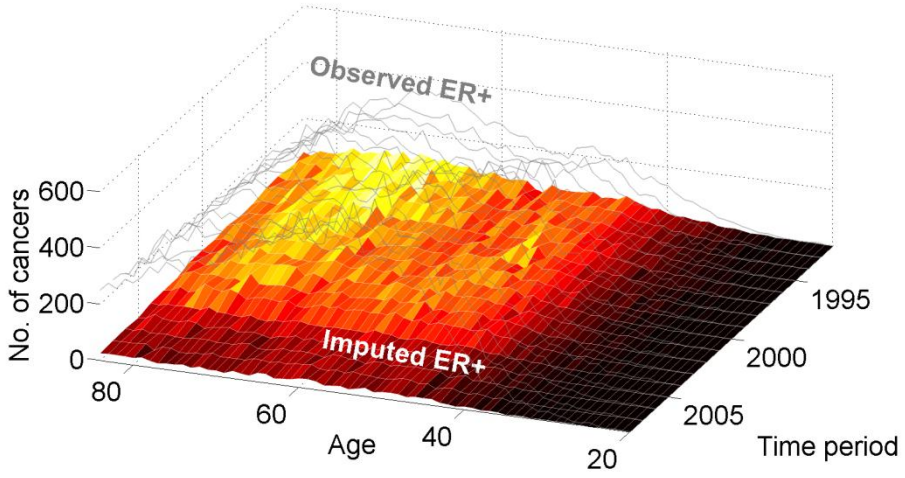
in which $\hat{\gamma}_{at}^{b} = (P_{at}^{b} + N_{at}^{b}) / I_{at}^{b}$ is the observed fraction of women with known ER status for each age a and calendar year t. Then summing $\hat{P}_{at}^{Cb}$ and $\hat{N}_{at}^{Cb}$ over all ages yields unbiased estimators of the true numbers of ER positive and ER negative breast cancers each year for that replicate. The usual estimates of age-standardized incidence and estimated annual percentage change were computed by the imputed breast cancer counts for each replicate. We estimated the variance of the $B = 2000$ bootstrap replicates for each of the estimated incidences, estimated annual percentage change, and estimated confidence intervals using the percentiles of the bootstrap distributions of each quantity (Supplementary Table 1).

We also developed an extended imputation model that incorporated the American Joint Committee on Cancer TNM stage (3) and tumor grade (Supplementary Table 2). Although assignments varied at the individual level, the overall imputed counts for this extended model were very similar to our basic model that was conditioned on age and year of diagnosis.

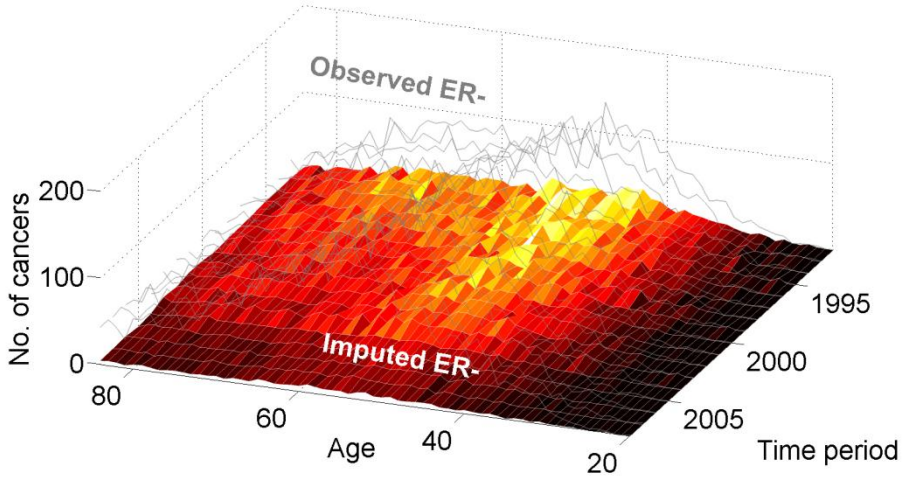
Finally, we used percentiles of the bootstrap distribution to estimate standard errors, confidence limits, and variance-covariance matrices for parameters in the age-period-cohort model (4, 5) (Supplementary Figure 3). For the breast cancer data in this study, adjustment for reporting delay and reporting error (6) had negligible impact on breast cancer trends and was not considered for further analysis.

Supplementary Figure 1

A

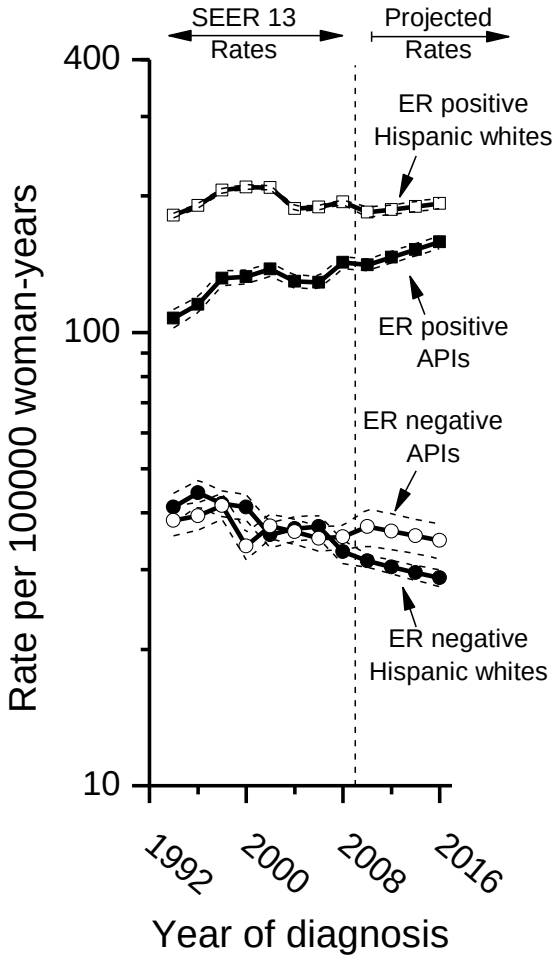


B



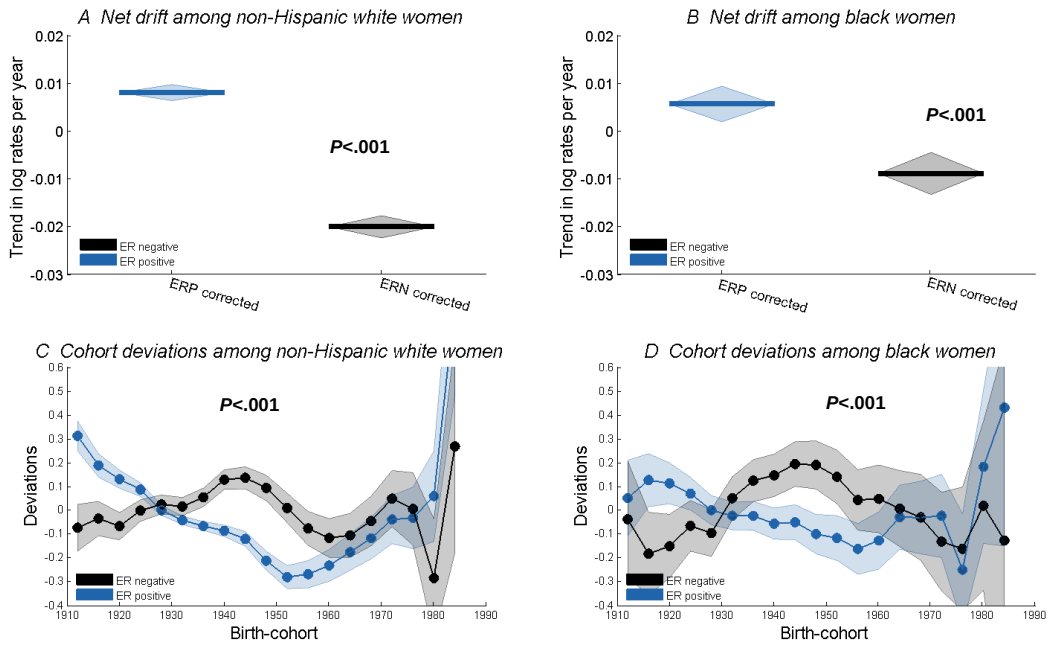
Supplementary Figure 1. Observed and imputed estrogen receptor (ER) positive and negative cancers in the National Cancer Institute's Surveillance, Epidemiology and End Results 13 Registries Database (1992 through 2008) (7) among women aged 20–84 years with invasive female breast cancer. **A**) Surface plot for imputed ER positive cancers are shown. Color maps show the lowest to highest counts from black to red to orange to yellow color scheme and are superimposed on a mesh plot of observed ER positive breast cancers by age at diagnosis and year of diagnosis. **B**) Surface and mesh plots for ER negative imputed and observed cancers are shown. Greater imputation of ER unknown cancers were required for earlier compared with recent years.

Supplementary Figure 2



Supplementary Figure 2. Imputed incidence rates of breast cancer by estrogen receptor (ER) status and race. The imputed incidence rates are shown for Hispanic whites and Asian or Pacific Islander (API) race/ethnicity. **Dashed lines** represent 95% confidence intervals. We used patient and population data from the Surveillance, Epidemiology and End Results 13 Registries Database (SEER 13) (7).

Supplementary Figure 3



Supplementary Figure 3. Age-period-cohort parameters for imputed estrogen receptor (ER) positive and negative net drifts and birth cohort deviations among non-Hispanic white and black women with invasive female breast cancers in the Surveillance, Epidemiology and End Results 13 Registries Database between 1992 through 2008 (7). Point estimates (**bars**) and 95% confidence intervals (**envelopes**) for the ER negative and ER positive net drifts among **A**) non-Hispanic white women and **B**) black women are shown. Point estimates (**solid circles**) and 95% confidence intervals (**envelopes**) for ER negative and ER positive birth-cohort deviations among **C**) non-Hispanic white women and **D**) black women are also shown. *P* values assess the null hypothesis of no difference between net drifts and/or birth cohort deviations within racial/ethnic groups. All statistical tests were two-sided.

Supplementary Table 1 Imputed (or corrected) annual percentage change in the age-standardized incidence rate by estrogen receptor expression, age, and racial/ethnic group for patients in the Surveillance, Epidemiology and End Results 13 Registries Database (1992 through 2008) *†

Population age	Age-standardized incidence rate of ER positive breast cancer		Age-standardized incidence rate of ER negative breast cancer	
	No. of cancers	% Change (95% CI)	No. of Cancers	% Change (95% CI)
All races combined, y				
30–84	312597	0.01 (-0.07 to 0.08)	90950	-1.71 (-1.85 to -1.57)
30–49	68015	1.17 (1.00 to 1.33)	30369	-2.42 (-2.66 to -2.18)
50–84	244582	-0.32 (-0.40 to -0.24)	60581	-1.35 (-1.52 to -1.19)
Non-Hispanic whites, y				
30–84	238516	0.17 (0.09 to 0.26)	60596	-1.95 (-2.12 to -1.79)
Blacks, y				
30–84	23012	0.45 (0.17 to 0.73)	13749	-0.93 (-1.30 to -0.56)

* CI = confidence interval, ER = estrogen receptor.

† ER positive and negative incidence trends were corrected for missing ER data conditioned on age and year of diagnosis.

Supplementary Table 2. Imputed (or corrected) annual percentage change in the age-standardized incidence rate by estrogen receptor expression, age, and racial/ethnic group for patients in the Surveillance, Epidemiology and End Results 13 Registries Database (1992 through 2008) *†

Population age	Age-standardized incidence rate of ER positive breast cancer		Age-standardized incidence rate of ER negative breast cancer	
	No. of cancers	% Change (95% CI)	No. of cancers	% Change (95% CI)
All races combined, y				
30–84	312420	-0.01 (-0.09 to 0.07)	91119	-1.66 (-1.81 to -1.52)
30–49	67970	1.15 (0.99 to 1.31)	30406	-2.38 (-2.62 to -2.13)
50–84	244450	-0.33 (-0.42 to -0.25)	60713	-1.30 (-1.47 to -1.13)
Non-Hispanic whites, y				
30–84	238480	0.16 (0.07 to 0.24)	60613	-1.90 (-2.07 to -1.72)
Blacks, y				
30–84	22950	0.47 (0.17 to 0.76)	13585	-0.79 (-1.17 to -0.41)

* CI = confidence interval, ER = estrogen receptor.

† ER positive and negative incidence trends were corrected for missing ER data conditioned on age and year of diagnosis, American Joint Committee on Cancer TNM stage (3), and tumor grade.

REFERENCES

1. Efron B, Tibshirani R. *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability 57*. New York: Chapman & Hall; 1993.

2. Cochran WG. *Sampling Techniques* 3rd ed. New York: John Wiley & Sons, Inc; 1977.

3. AJCC. Breast. In: Greene FL, Page DL, Fleming ID, et al., eds. *AJCC Cancer Staging Handbook*. 6 ed. New York: Springer; 2002:255-281.

4. Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Stat. Med.* Jun 1987;6(4):449-467.

5. Holford TR. Age-Period-Cohort Analysis. In: Armitage P, Colton T, eds. *Encyclopedia of Biostatistics.* Vol 1. Second Edition ed. West Sussex: John Wiley & Sons Ltd; 2005:105-123.

6. Clegg LX, Feuer EJ, Midthune DN, Fay MP, Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. *J Natl Cancer Inst.* Oct 16 2002;94(20):1537-1545.

7. SEER-13. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence-SEER 13 Regs Research Data, Nov 2010 Sub (1992-2008) <Single Ages to 85+, Katrinia/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969-2009 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2011, based on November 2010 submission. 2011; http://seer.cancer.gov/. Accessed May 20, 2011.