Annals of Internal Medicine ORIGINAL RESEARCH

Estimating Breast Cancer Overdiagnosis After Screening Mammography Among Older Women in the United States

Ilana B. Richman, MD, MHS; Jessica B. Long, MPH; Pamela R. Soulos, MPH; Shi-Yi Wang, MD, PhD; and Cary P. Gross, MD

Background: Overdiagnosis is increasingly recognized as a harm of breast cancer screening, particularly for older women.

Objective: To estimate overdiagnosis associated with breast cancer screening among older women by age.

Design: Retrospective cohort study comparing the cumulative incidence of breast cancer among older women who continued screening in the next interval with those who did not. Analyses used competing risk models, stratified by age.

Setting: Fee-for-service Medicare claims, linked to the SEER (Surveillance, Epidemiology, and End Results) program.

Patients: Women 70 years and older who had been recently screened.

Measurements: Breast cancer diagnoses and breast cancer death for up to 15 years of follow-up.

Results: This study included 54 635 women. Among women aged 70 to 74 years, the adjusted cumulative incidence of breast cancer was 6.1 cases (95% CI, 5.7 to 6.4) per 100 screened women versus 4.2 cases (CI, 3.5 to 5.0) per 100 unscreened women. An estimated 31% of breast cancer among screened women were potentially overdiagnosed. For women aged 75 to 84 years, cumulative incidence was 4.9 (CI, 4.6 to 5.2) per 100 screened women versus 2.6 (CI, 2.2 to 3.0) per 100 unscreened women, with 47% of cases potentially overdiagnosed. For women aged 85 and older, the cumulative incidence was 2.8 (CI, 2.3 to 3.4) among screened women versus 1.3 (CI, 0.9 to 1.9) among those not, with up to 54% overdiagnosis. We did not see statistically significant reductions in breast cancer–specific death associated with screening.

Limitations: This study was designed to estimate overdiagnosis, limiting our ability to draw conclusions on all benefits and harms of screening. Unmeasured differences in risk for breast cancer and differential competing mortality between screened and unscreened women may confound results. Results were sensitive to model specifications and definition of a screening mammogram.

Conclusion: Continued breast cancer screening was associated with greater incidence of breast cancer, suggesting overdiagnosis may be common among older women who are diagnosed with breast cancer after screening. Whether harms of overdiagnosis are balanced by benefits and for whom remains an important question.

Primary Funding Source: National Cancer Institute.

Ann Intern Med. doi:10.7326/M23-0133 **Annals.org**
For author, article, and disclosure information, see end of text.
This article was published at Annals.org on 8 August 2023.

Although older women are commonly screened for breast cancer, the efficacy of screening in this population remains uncertain (1). No randomized trials have evaluated screening mammography in women 75 years and older, and only a few studies have included women over the age of 70 years, leaving uncertainty about benefits and harms of screening in older women (2, 3). Observational studies suggest that the mortality benefit from screening may be limited to women younger than 75 years (4). Modeling studies, by contrast, indicate that screening reduces breast cancer mortality, but the net benefit of screening diminishes with increasing age and comorbidity (5, 6). Guidelines about screening older women vary. The U.S. Preventive Services Task Force makes no specific recommendation for or against screening women 75 years and older, but includes women 70 to 74 years in the broader group of women for whom screening is generally recommended (7). The American Cancer Society recommends continuing screening if life expectancy is more than 10 years, whereas the American College of Physicians recommends discontinuing screening at age 75 years or younger if life expectancy is less than 10 years (8, 9).

Harms of screening for older women include frequent false positives requiring additional testing and invasive procedures (10–12). However, in recent years, there has also been greater recognition that overdiagnosis constitutes an important harm from breast cancer screening. Overdiagnosis may be defined as detecting a cancer, often through screening, that would not have caused symptoms in a person's lifetime (13). Risk for overdiagnosis is driven by several factors, including the biological behavior of a tumor and life expectancy (14). Specifically, some types of breast cancer may have a long presymptomatic phase. Detecting these types of breast cancer through screening may result in overdiagnosis if these types of breast cancer would have otherwise remained clinically silent during a patient's lifetime. In addition, even aggressive types of breast cancer with a short presymptomatic phase may be overdiagnosed in older women who have very limited life expectancy. Modeling studies have estimated that overdiagnosis

See also:

Editorial comment

Web-Only
Supplement

may occur in approximately 0.2 to 7.5 older women per 1000 screened for breast cancer depending on age, comorbidity, and the specific model, and may account for between 12% and 48% of screen-detected breast cancer (5, 15). Although important, modeling studies have some inherent limitations. For example, modeling studies make assumptions about the distribution of lead times of breast cancer, which are not directly observable (16, 17). Studies using empirical or observational methods have often focused on younger screened populations or all screened women, rather than older women specifically (18, 19).

The primary goal of this study was to quantify the risk for overdiagnosis associated with screening mammography among older women by evaluating the difference in cumulative incidence of breast cancer associated with continuing screening or not in the next scheduled interval. To do this, we approximated a target trial in which women 70 years and older who had recently been screened and did not have a history of breast cancer would be assigned to either continue screening for at least 1 more round or not at the time of their next mammogram. We stratified analyses by age (70 to 74 years, 75 to 84 years, and 85 years and older). Because some screening recommendations use life expectancy instead of age, and because life expectancy can vary within age groups, we replicated our analyses using life-expectancy strata.

METHODS

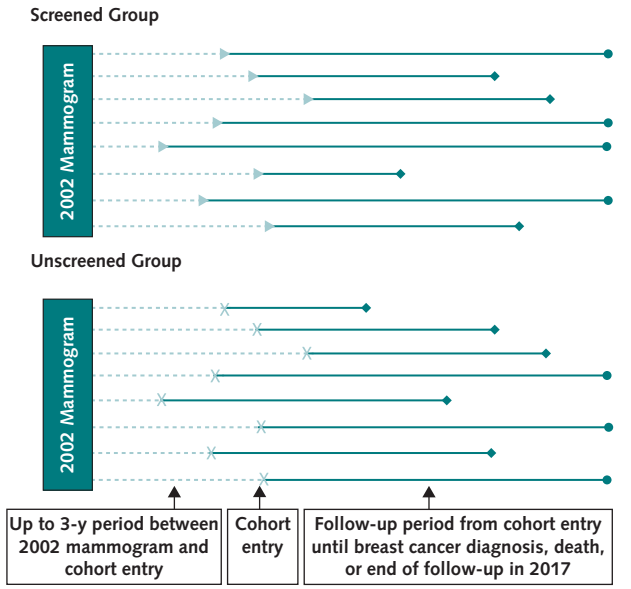
Data

We used data from the SEER (Surveillance, Epidemiology, and End Results)-Medicare registry linked to a 5% sample of Medicare fee-for-service beneficiaries (20, 21). This sample includes Medicare beneficiaries who were ultimately diagnosed with breast cancer, those diagnosed with other types of cancer, and those who were not diagnosed with cancer. Follow-up was available through 2017 (22).

Target Trial and Cohort Selection

This study was designed to approximate a target trial of the effect of having another screening mammogram during the next interval or not on cumulative incidence of breast cancer among women aged 70 years and older who had been recently screened and who did not have a history of breast cancer. To implement this, we first identified women who had been screened in 2002, were aged 70 years or older by 1 January 2003, had not had breast cancer before their 2002 screening mammogram, and had Medicare fee-for-service insurance through 2005. We then considered a 3-year period after the 2002 mammogram during which women could either be screened or not. For women who continued screening during this period, cohort entry (which may be thought of as "time zero") began on the day of the next screening mammogram. For women who did not have a screening mammogram within 3 years of their 2002 mammogram, we assigned a "pseudomammogram" date, meant to represent the date on which a screening mammogram would have occurred, had that woman been screened.

Figure 1. Study schematic.



Each horizontal line represents a person in the study. Study entry begins at the date of the mammogram (▶) or pseudomammogram (X), which must be within 3 years of the 2002 mammogram. The time between the 2002 mammogram and study entry is similar for both the screened and unscreened groups, and both groups include only women who have survived and are breast cancer–free at the time of cohort entry. The solid bars represent the follow-up period, which begins at the time of the mammogram or pseudommamogram date and ends either at death or breast cancer diagnosis (♦) or end of follow-up in 2017 (●).

The pseudomammogram date was chosen at random from the distribution of times to the next mammogram among women who were screened. Women were excluded from this nonscreening group if they received a nonscreening mammogram, were diagnosed with breast cancer, or died before the pseudomammogram date, but not after. Therefore, women in both groups (screened and unscreened) had survived and were free of breast cancer between their 2002 mammogram until cohort entry, which was the day of their next screening mammogram or the pseudomammogram date (**Figure 1**) (23).

Exposure Definition

We defined screening mammography in Medicare claims using an algorithm developed by Fenton and colleagues that distinguishes screening mammograms from diagnostic mammograms in claims data (**Supplement Methods** and **Supplement Table 1**, both available at Annals.org) (24). The algorithm has a sensitivity of 99.7% and a specificity of 62.7% but maintains a high positive predictive value for identifying screening mammograms (97.4%) because most mammograms performed are screening mammograms. We used this approach to identify women who underwent screening mammography in the 3 years after their 2002 mammogram. Women who did not have a screening mammogram during this time frame were included in the nonscreening group, as described.

Outcome Definition

The primary outcome in this study was breast cancer diagnosis, as captured in the SEER registry. We included all breast cancer diagnoses including in situ carcinomas. We also evaluated use of screening mammography over time in each group. Secondary outcomes included breast cancer diagnosis by stage (in situ, localized invasive, and regional–distant stage) based on SEER summary stage, a variable available across the long range of follow-up for all registries included in the sample. Lastly, we evaluated breast cancer–specific mortality, as documented by SEER using death certificate records.

Covariates

We evaluated demographic and clinical characteristics of the cohort including age, race, ethnicity, urban–rural status, state buy-in, ZIP code poverty, receipt of flu vaccine, and frailty. State buy-in indicates state payment for Medicare premiums and approximates Medicare–Medicaid dual eligibility. Frailty was defined using the Kim index, dichotomized at a value of 0.2 (25). We calculated life expectancy for each individual using age, sex, and comorbidity at the cohort entry date using an established method (26).

Analysis

We compared characteristics of the study population by screening status, calculating standardized mean differences to evaluate differences between screened and unscreened women within age groups. We also evaluated patterns of screening after cohort entry by calculating the proportion of women screened at subsequent 3-year intervals after cohort entry by age group and by screening status at cohort entry.

To estimate overdiagnosis, we compared the cumulative incidence of breast cancer among women screened at cohort entry to that among women not screened at cohort entry. To calculate cumulative incidence, we fit a competing risk model using the Fine–Gray method, accounting for the competing risk for death (27). This approach allows for the estimation of cumulative incidence of breast cancer when competing events that preclude the possibility of breast cancer diagnosis (like death from other causes) are common. Models were stratified by age at cohort entry (70 to 74 years, 75 to 84 years, ≥85 years) or life expectancy at cohort entry (≤5 years, 6 to 10 years, >10 years). We adjusted models for variables that may influence both screening use and the underlying risk for breast cancer, specifically age, race, and ethnicity. We also adjusted models for factors that may influence both screening use and competing risk for mortality, specifically life expectancy (continuous in months), frailty, state buy-in, and receipt of a flu shot, which may be more common among those who are healthier and also seek out preventive care (28, 29).

We estimated the cumulative incidence of breast cancer for screened and unscreened women at the end of follow-up using mean values for the population in each age group. As our main measure of overdiagnosis, we calculated the absolute risk difference, which we defined as the difference in the cumulative incidence of breast cancer among women who were screened versus not screened at cohort entry. We used a bootstrap approach with 1000 replicate samples to estimate 95% CIs for our estimates (30). Lastly, we quantified the risk for overdiagnosis among screened women diagnosed with breast cancer. We defined this as the absolute risk difference (difference in cumulative incidence of breast cancer between screened and unscreened women) divided by the cumulative incidence among screened women. This quantity reflects the proportion of breast cancer cases among screened women that may be overdiagnosed. Our approach for stage-specific incidence was identical, except we used stage-specific breast cancer diagnosis as the primary outcome, with breast cancer diagnosis at other stages as a competing event. For breast cancer mortality analyses, we used the same approach as our main analysis, but calculated cumulative incidence of breast cancer death at the end of follow-up rather than breast cancer incidence.

Sensitivity Analyses

Identifying screening mammograms relies on a claims-based algorithm that distinguishes screening and diagnostic mammograms. This algorithm in general classifies the great majority of screening mammograms correctly, with less than 2% of mammograms incorrectly classified as screening when they are actually diagnostic. However, because even this small misclassification may impact results, we conducted a sensitivity analysis in which we conservatively favored categorizing women as not screened when misclassification was possible (**Supplement Methods**). We also evaluated the rate of cancer diagnosis within 12 months of mammograms reclassified under this alternate definition to understand whether diagnostic yield was similar to screening mammograms or not.

In addition to our primary analyses, we tested alternate model specifications. We fit cause-specific hazard models in addition to Fine–Gray models. Cause-specific hazard models are less susceptible to confounding from competing events, but may overestimate cumulative incidence (28). We also used logistic models, estimating the predicted probability of breast cancer diagnosis at 15 years for women who were screened or not screened by life expectancy to investigate potential model sensitivity to violation of the proportional hazards assumption. We performed a sensitivity analysis in which we censored women if they received a screening mammogram more than 8 years after cohort entry to ensure sufficient follow-up time to observe breast cancer diagnoses (31). Lastly, we evaluated the potential effect of family history as an unmeasured confounder on our results (**Supplement Methods**).

Role of the Funding Source

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health. The funders had no role in the design, conduct, or reporting of this research.

RESULTS

The cohort included 54 635 women (**Table 1** and **Supplement Table 2**). The mean age of the population

Table 1. Descriptive Characteristics of Study Sample

Characteristic	Age 70–74 y			Age 75–84 y			Age ≥85 y		
	Screened	Unscreened	Standardized Difference	Screened	Unscreened	Standardized Difference	Screened	Unscreened	Standardized Difference
Beneficiaries, *n*	17 488	2437	–	23 613	5707	–	3384	2006	–
Mean age (95% CI), *y*	72.0 (72.0–72.0)	72.1 (72.0–72.1)	0.04	78.6 (78.6–78.7)	79.3 (79.2–79.3)	0.23*	87.2 (87.1–87.3)	87.8 (87.7–87.9)	0.27*
Mean LE† (95% CI), *mo*	118.3 (118.1–118.4)	115.6 (115.0–116.2)	0.22*	105.6 (105.4–105.9)	97.7 (97.0–98.4)	0.33*	62.4 (61.7–63.0)	56.9 (56.0–57.8)	0.28*
LE ≤5 y, *n (%)*	116 (1)	62 (3)	0.15*	1263 (5)	666 (12)	0.23*	1418 (42)	1069 (53)	0.23*
LE 6–10 y, *n (%)*	1055 (6)	280 (11)	0.19*	10 484 (44)	2976 (52)	0.16*	1966 (58)	937 (47)	0.23*
LE >10 y, *n (%)*	16 317 (93)	2095 (86)	0.24*	11 866 (50)	2065 (36)	0.29*	NA	NA	NA
Race and ethnicity,‡ *n (%)*									
Black	1005 (6)	216 (9)	0.12*	1130 (5)	388 (7)	0.09	159 (5)	113 (6)	0.04
Other	618 (4)	134 (5)	0.09	626 (3)	212 (4)	0.06	84 (2)	47 (2)	0.01
White	15 184 (87)	1936 (79)	0.20*	21 258 (90)	4846 (85)	0.15*	3070 (91)	1796 (90)	0.04
Hispanic	681 (4)	151 (6)	0.11*	599 (3)	261 (5)	0.11*	71 (2)	50 (2)	0.03
Comorbidity,§ *n (%)*									
0	10 533 (60)	1186 (49)	0.23*	12 723 (54)	2494 (44)	0.20*	1547 (46)	819 (41)	0.10*
1–2	5961 (34)	989 (41)	0.13*	8948 (38)	2373 (42)	0.08	1452 (43)	861 (43)	0.0003
≥3	994 (6)	262 (11)	0.19*	1942 (8)	840 (15)	0.20*	385 (11)	326 (16)	0.14*
Flu vaccine in prior 12 mo, *n (%)*	10 322 (59)	1152 (47)	0.24*	14 791 (63)	3132 (55)	0.16*	2190 (65)	1176 (59)	0.13*
PCP visit in prior 12 mo, *n (%)*	14 451 (83)	1944 (80)	0.07	19 716 (83)	4612 (81)	0.07	2790 (82)	1626 (81)	0.04
Frail‖	1635 (9)	453 (19)	0.27*	3219 (14)	1443 (25)	0.30*	681 (20)	605 (30)	0.23*
ZIP code–level poverty, *n (%)*									
<5%	4073 (23)	450 (18)	0.12*	5597 (24)	1207 (21)	0.06	734 (22)	444 (22)	0.01
5% to 9.9%	5433 (31)	703 (29)	0.05	7496 (32)	1678 (29)	0.05	1106 (33)	586 (29)	0.08
10% to 19.9%	5321 (30)	814 (33)	0.06	7046 (30)	1753 (31)	0.02	1005 (30)	595 (30)	0.001
≥20%	2211 (13)	394 (16)	0.10*	2858 (12)	868 (15)	0.09	445 (13)	300 (15)	0.05
Unknown	450 (3)	76 (3)	0.03	616 (3)	201 (4)	0.05	94 (3)	81 (4)	0.07
State buy-in,¶ *n (%)*	1538 (9)	460 (19)	0.30*	1761 (7)	904 (16)	0.26*	269 (8)	245 (12)	0.14*
Nonmetro residence,** *n (%)*	2942 (17)	441 (18)	0.03	3559 (15)	942 (17)	0.04	473 (14)	289 (14)	0.01

LE = life expectancy; NA = not applicable; PCP = primary care provider.
* Values denoted with an asterisk indicate that the standardized mean difference is ≥0.1.
† Life expectancy (LE) was calculated from age, sex, and comorbidities using the method by Tan et al (26).
‡ For race and ethnicity, the "other" category includes the following groups, which were combined to maintain privacy because of small cell size: Asian, North American Native, Other, Unknown.
§ Comorbidity categories indicate number of Elixhauser conditions previously found to be significantly associated with reduced survival in a non-cancer cohort.
‖ Frailty was calculated from procedure and diagnosis codes using an algorithm by Kim et al (25), dichotomized at a score of 0.2.
¶ State buy-in refers to patients for whom the state pays Medicare premiums, an approximation of dual Medicare/Medicaid eligibility.
** Nonmetro residence was defined using state and county in 2003 with Rural Urban Continuum Codes.

was 77.2 years (95% CI, 77.1 to 77.2 years), 6% of women were Black, 3% were Hispanic, and 88% were White. Life expectancy was 10 years or less for 41% of the population and 15% were considered frail. Across age groups, women who underwent screening had longer life expectancy and were less likely to have state buy-in or to be considered frail (**Table 1**).

Among women aged 70 to 74 years, 88% were screened at cohort entry (that is, within 3 years of the 2002 mammogram). Among women aged 75 to 84 years, 81% were screened at cohort entry, and among women aged 85 years and older, 63% were screened at cohort entry (**Table 1**). In all age categories, some women who were not screened at cohort entry were screened during a later time interval. Among women 70 to 74 years who were not screened at cohort entry, 30% were screened in the first 3 years of follow-up (**Supplement Figure, *A***, available at Annals.org). Among women with a life expectancy of 75 to 84 years not screened at cohort entry, 16% were screened in the first 3 years (**Supplement Figure, *B***). For women aged 85 years and older, 6% were screened in the first 3 years of follow-up (**Supplement Figure, *C***).

Median follow-up times were 13.7 years (IQR, 9.2 to 14.4 years) among women aged 70 to 74 years, 10 years (IQR, 5.8 to 13.9 years) for women aged 75 to 84 years, and 5.7 years (IQR, 3.1 to 9.1 years) for women 85 years and older. By the end of follow-up, among those 70 to 74 years who were screened, 38% had died, versus 56% among those who were not screened. Among those aged 75 to 84 years, 65% of those who were screened

had died versus 80% among those who were not screened. For those aged 85 years and older, 91% of those screened had died versus 96% among those not screened.

In adjusted analyses using Fine–Gray competing risk models, the cumulative incidence of breast cancer was 6.1 cases (CI, 5.7 to 6.4 cases) per 100 women among those 70 to 74 years who were screened at cohort entry, versus 4.2 cases (CI, 3.5 to 5.0 cases) per 100 women among those who were not screened at cohort entry (risk difference, 1.9 cases [CI, 1.0 to 2.8 cases] per 100) (**Figure 2** and **Table 2**). Among women screened at cohort entry who were eventually diagnosed with breast cancer, we estimated 31% may be overdiagnosed. Among women aged 75 to 84 years, the cumulative incidence of breast cancer was 4.9 (CI, 4.6 to 5.2) per 100 among women who were screened at cohort entry versus 2.6 (CI, 2.2 to 3.0) per 100 among women who were not screened at cohort entry (risk difference, 2.3 [CI, 1.7 to 2.8]) (**Figure 2** and **Table 2**). We estimated that 47% of breast cancer cases among screened women may be overdiagnosed. For women 85 years and older who were screened, cumulative incidence of breast cancer was 2.8 (CI, 2.3 to 3.4) per 100 versus 1.3 (CI, 0.9 to 1.9) per 100 among women not screened at cohort entry (risk difference, 1.5 [CI, 0.6 to 2.2]) (**Figure 2** and **Table 2**). Risk for overdiagnoses was estimated at 54% among screened women diagnosed with breast cancer. When stratifying by life expectancy, an estimated 32% of breast cancer among screened women with a life expectancy of more than 10 years was overdiagnosed. Among women with a life expectancy of 6 to 10 years, 53% of cancer was potentially overdiagnosed, and 63% among women with a life expectancy of 5 years or fewer (**Supplement Table 3**, available at Annals.org).

In sensitivity analyses, models using logistic regression, and models that censored women if screening was performed more than 8 years after cohort entry generated similar estimates of risk difference. Estimates of overdiagnosis from cause-specific hazard models were somewhat lower than estimates from Fine-Gray models (**Table 2**). Findings were also sensitive to the definition of a screening mammogram. Using an alternate, more conservative definition of a screening mammogram, the risk difference between screened and unscreened women among those 70 to 74 years was 0.9 breast cancer cases (CI, −0.1 to 1.7) per 100, with an estimated 15% of screen-detected cancer overdiagnosed. For women aged 75 to 84 years, the risk difference was 1.7 (CI, 1.1 to 2.2) with an estimated 36% of breast cancer cases overdiagnosed. For women aged 85 years and older, the risk difference was 1.1 (CI, 0.3 to 1.7) after 15 years of follow-up with an estimated 44% of screen-detected cancer overdiagnosed (**Table 2**). Breast cancer diagnosis was more common among mammograms reclassified as not screening in this sensitivity analysis, suggesting that some of these mammograms may have been diagnostic (**Supplement Table 4**, available at Annals.org). Estimates of the effect of family history suggested that differential screening use among women with a first-degree relative with breast cancer would not explain our results (**Supplement Table 5**, available at Annals.org).

Figure 2. Cumulative incidence of breast cancer by screening status and age.

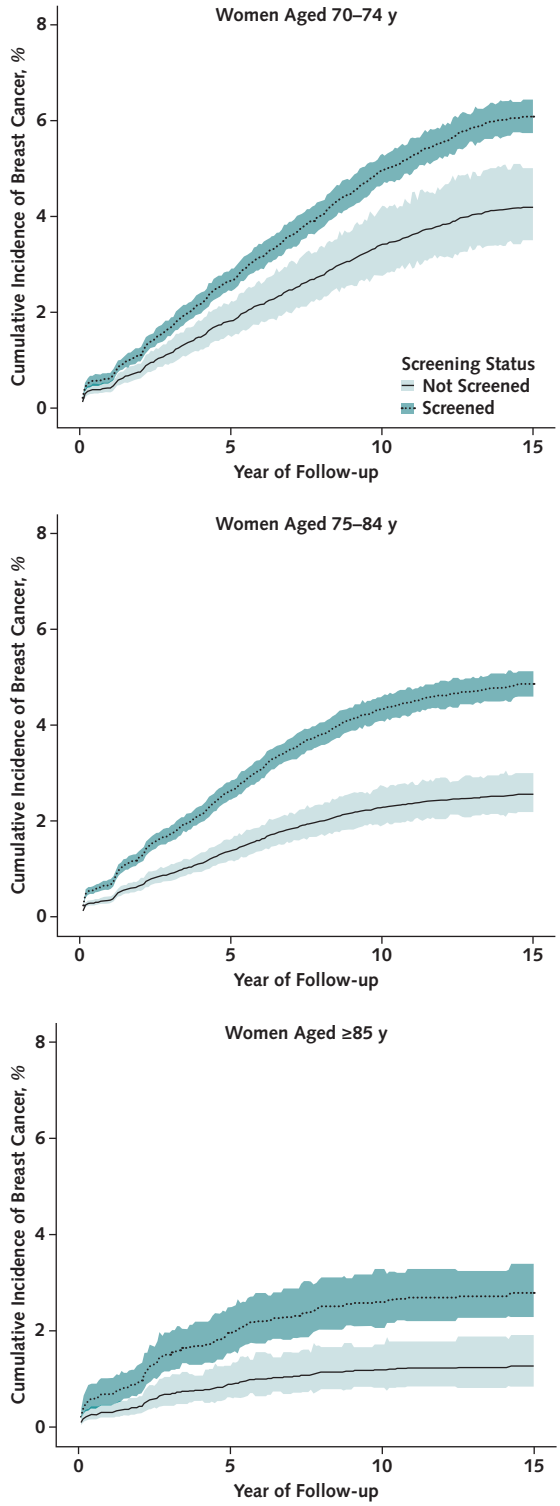


Figure panels depict cumulative incidence of breast cancer (breast cancer cases per 100 women) among women screened or not screened at cohort entry over available follow-up. Shaded areas indicate 95% CIs. *Top*. Age 70 to 74 years. *Middle*. Age 75 to 84 years. *Bottom*. Age 85 years and older.

Table 2. Cumulative Incidence of Breast Cancer Cases per 100 Persons*

Exposure	Primary Analysis†		Sensitivity Analyses†‡			
	Unadjusted	Adjusted	Cause-Specific	Logistic Regression	Censored if Screened >8 y After Cohort Entry	Alternate Screening Definition§
Age 70–74 y						
Not screened	4.0 (3.3 to 4.9)	4.2 (3.5 to 5.0)	5.5 (4.6 to 6.6)	4.2 (3.4 to 5.1)	3.6 (2.9 to 4.5)	4.9 (4.2 to 5.8)
Screened	6.2 (5.9 to 6.6)	6.1 (5.7 to 6.4)	7.1 (6.6 to 7.5)	6.0 (5.7 to 6.4)	5.1 (4.8 to 5.5)	5.8 (5.5 to 6.2)
Difference	2.2 (1.3 to 3.0)	1.9 (1.0 to 2.8)	1.6 (0.4 to 2.7)	1.9 (1.0 to 2.7)	1.5 (0.6 to 2.3)	0.9 (−0.1 to 1.7)
Excess, %	35	31	22	31	29	15
Hazard ratio‖	1.56 (1.27 to 1.91)	1.47 (1.19 to 1.81)	1.29 (1.05 to 1.59)	1.48 (1.20 to 1.83)	1.41 (1.12 to 1.78)	1.19 (0.99 to 1.43)
Age 75–84 y						
Not screened	2.4 (2.1 to 2.8)	2.6 (2.2 to 3.0)	4.1 (3.4 to 4.8)	2.6 (2.2 to 3.0)	2.3 (1.9 to 2.7)	3.0 (2.7 to 3.5)
Screened	5.0 (4.8 to 5.3)	4.9 (4.6 to 5.2)	6.4 (6.0 to 6.8)	4.8 (4.5 to 5.1)	4.4 (4.1 to 4.6)	4.7 (4.4 to 5.0)
Difference	2.6 (2.1 to 3.1)	2.3 (1.7 to 2.8)	2.3 (1.5 to 3.1)	2.3 (1.7 to 2.8)	2.1 (1.6 to 2.6)	1.7 (1.1 to 2.2)
Excess, %	52	47	36	47	47	36
Hazard ratio‖	2.10 (1.76 to 2.50)	1.92 (1.60 to 2.30)	1.59 (1.33 to 1.91)	1.93 (1.61 to 2.31)	1.93 (1.59 to 2.33)	1.56 (1.32 to 1.83)
Age ≥85 y						
Not screened	1.3 (0.9 to 2)	1.3 (0.9 to 1.9)	3.2 (2.0 to 5.1)	1.3 (0.9 to 1.9)	1.3 (0.9 to 1.9)	1.4 (1.0 to 2.0)
Screened	2.9 (2.3 to 3.7)	2.8 (2.3 to 3.4)	5.6 (4.2 to 7.5)	2.8 (2.3 to 3.4)	2.7 (2.2 to 3.4)	2.5 (1.9 to 3.1)
Difference	1.6 (0.8 to 2.4)	1.5 (0.6 to 2.2)	2.4 (0.6 to 4.2)	1.5 (0.6 to 2.2)	1.4 (0.6 to 2.0)	1.1 (0.3 to 1.7)
Excess, %	55	54	43	53	52	44
Hazard ratio‖	2.56 (1.46 to 3.47)	2.20 (1.43 to 3.40)	1.78 (1.15 to 2.76)	2.15 (1.39 to 3.33)	2.13 (1.37 to 3.29)	1.76 (1.15 to 2.69)

* The table presents the cumulative incidence of breast cancer (breast cancer cases per 100 persons) at the end of follow-up, which occurred at death, breast cancer diagnosis, or through the end of 2017. All models use the Fine-Gray method, except the logistic model and the cause-specific hazard model.
† Values indicate cumulative incidence (95% CI).
‡ All sensitivity analyses used the same set of covariates as in the primary adjusted analysis.
§ The alternate screening definition reclassifies women who received mammograms billed with diagnostic codes in the absence of claims for breast cancer symptoms as "not screened" rather than "screened."
‖ Hazard ratios compare risk for breast cancer diagnosis in screened and unscreened groups. Logistic models produce odds ratios rather than hazard ratios.

Lastly, we evaluated secondary outcomes including cumulative incidence by stage (in situ, localized invasive, and regional or distant breast cancer) and breast cancer–specific mortality. Cumulative incidence was higher among screened women both for in situ breast cancer and localized invasive cancer across age groups (Table 3). We did not see statistically significant higher or lower incidence of regional–distant breast cancer by screening status. We also did not see statistically significant differences in breast cancer–specific mortality by screening status (Table 3).

DISCUSSION

We found that the proportion of breast cancer that may be overdiagnosed among older women who are screened is considerable, and increases with advancing age and with decreasing life expectancy. For women 85 years and older, 54% of breast cancer among screened women may be overdiagnosed. For younger women, aged 70 to 74 years, the proportion is smaller but still considerable at up to 31%. We also saw that the absolute risk for overdiagnosis was similar across age groups and ranged from 1.5 to 2.3 cases per 100 women screened. The higher proportion of overdiagnosed cases among older women reflects the fact that although the absolute risk is similar across age groups, the cumulative incidence of breast cancer is lower among older women who have greater competing mortality.

Is an absolute risk for overdiagnosis of about 2% after 15 years high? Whether this risk is considered high depends on several factors including expected benefits of screening and patient preferences. We evaluated the association between breast cancer screening and breast cancer–specific death to understand potential benefits of screening in this population. Although we did not see statistically significant reductions in death from breast cancer in any age or life-expectancy stratum, point estimates suggested reduction in breast cancer–specific death for women younger than 85 years, consistent with some modeling studies (5, 6). However, uncertainty around our estimates precludes drawing strong conclusions about mortality benefits in this analysis, and other observational studies have documented no mortality benefit for screening among women older than 75 years (4). Given uncertainty about the relative balance of benefits and harms of screening in this population, patient preferences, including risk tolerance, comfort with uncertainty, and willingness to undergo treatment, are important for informing screening decisions.

Stage-specific analyses suggested overdiagnosis was driven by in situ and localized invasive breast cancer rather than advanced breast cancer. Whether overdiagnosis of these types of early-stage cancer is consequential in part depends on whether diagnosis results in aggressive or burdensome treatments. Up to 90% of women aged 80 years and older with nonmetastatic breast cancer undergo surgery, and nearly two thirds of women older than age 70 years have radiation for early-stage invasive breast cancer (32, 33). Not only are these treatments intensive, but older women also risk functional decline after surgery (34). Importantly, some studies also suggest that continued

Table 3. Adjusted Cumulative Incidence of Stage-Specific Breast Cancer and Breast Cancer Death per 100 Persons*

Exposure	Overall Breast Cancer Incidence†	In Situ Breast Cancer Incidence†‡§	Localized Invasive Breast Cancer Incidence†‡	Regional–Distant Breast Cancer Incidence†‡	Breast Cancer Mortality†‖
Age 70–74 y					
Not screened	4.19 (3.49 to 5.03)	0.59 (0.36 to 0.98)	2.56 (2.05 to 3.20)	0.90 (0.61 to 1.34)	0.41 (0.22 to 0.76)
Screened	6.08 (5.74 to 6.44)	1.09 (0.94 to 1.27)	3.84 (3.58 to 4.11)	1.00 (0.85 to 1.17)	0.35 (0.26 to 0.48)
Difference	1.89 (0.98 to 2.75)	0.50 (0.10 to 0.81)	1.28 (0.51 to 1.93)	0.10 (−0.31 to 0.47)	−0.06 (−0.34 to 0.16)
Excess, %	31	46	33	10	−17
Hazard ratio¶	1.47 (1.19 to 1.81)	1.86 (1.08 to 3.18)	1.51 (1.15 to 1.98)	1.11 (0.71 to 1.72)	0.86 (0.44 to 1.68)
Age 75–84 y					
Not screened	2.56 (2.20 to 2.97)	0.15 (0.07 to 0.29)	1.50 (1.21 to 1.86)	0.74 (0.55 to 1.00)	0.42 (0.28 to 0.64)
Screened	4.85 (4.57 to 5.15)	0.79 (0.68 to 0.93)	3.15 (2.95 to 3.38)	0.78 (0.66 to 0.92)	0.36 (0.29 to 0.46)
Difference	2.29 (1.74 to 2.81)	0.64 (0.46 to 0.79)	1.65 (1.21 to 2.03)	0.04 (−0.21 to 0.27)	−0.06 (−0.27 to 0.11)
Excess, %	47	81	52	5	−17
Hazard ratio¶	1.92 (1.60 to 2.30)	5.41 (2.65 to 11.06)	2.12 (1.68 to 2.67)	1.05 (0.75 to 1.49)	0.87 (0.55 to 1.37)
Age ≥85 y					
Not screened	1.28 (0.87 to 1.89)	0.05 (0.01 to 0.21)	0.71 (0.39 to 1.29)	0.18 (0.08 to 0.38)	0.16 (0.05 to 0.50)
Screened	2.80 (2.30 to 3.41)	0.19 (0.11 to 0.34)	1.66 (1.26 to 2.20)	0.33 (0.17 to 0.64)	0.21 (0.09 to 0.51)
Difference	1.52 (0.65 to 2.20)	0.140 (0.001 to 0.220)	0.95 (0.29 to 1.38)	0.15 (−0.04 to 0.30)	0.05 (−0.12 to 0.19)
Excess, %	54	74	57	45	24
Hazard ratio¶	2.20 (1.43 to 3.40)	3.95 (0.98 to 15.97)	2.35 (1.32 to 4.19)	1.87 (0.69 to 5.12)	1.34 (0.4 to 4.49)

* The table presents the adjusted cumulative incidence (breast cancer cases or breast cancer deaths per 100 persons) through the end of follow-up, which occurred at breast cancer diagnosis, death, or the end of 2017.
† Values indicate cumulative incidence (95% CI).
‡ In situ, localized, and regional–distance incidence were derived from the SEER (Surveillance, Epidemiology, and End Results) summary stage variable.
§ For the in situ outcome, life expectancy was recoded into 6-month increments due to small cell sizes.
‖ Breast cancer death was identified from cause of death reported on death certificates. For the breast cancer death outcome, life expectancy was recoded into 6-month increments and race was analyzed as non-Hispanic Black compared with all others due to small cell sizes.
¶ Hazard ratios compare risk for breast cancer diagnosis or death in screened and unscreened groups.

screening is associated with lower rates of chemotherapy use, which is an important potential benefit of screening that must be weighed against the risks for overtreatment (4). Even beyond the specific burdens of treatment, the experience of being diagnosed with breast cancer is deeply affecting for many women and is associated with anxiety, reductions in quality of life, and lower sense of well-being (35).

Our findings are generally consistent with estimates from other studies. First, a recent study estimated that the sojourn time for progressive breast cancer is about 7 years (31). Given this, more than half of breast cancer identified among women with a mean life expectancy less than 7 years would be likely to be overdiagnosed. Indeed, we found that more than half (63%) of cases among women with a life expectancy of 5 years or less may be overdiagnosed and 54% of cases among women aged 85 years or more may be overdiagnosed. Our main results were somewhat higher than a modeling study that estimated that between 12% and 48% of screen-detected cancer among women 75 years and older are overdiagnosed, although those findings incorporate specific assumptions about the natural history of breast cancer (15). Lastly, our results echo findings that inferred overdiagnosis rates based on patterns of tumor size at diagnosis, and estimated that about half of breast cancer among women over the age of 80 years is overdiagnosed (36). Our work builds on this literature by using an approach that makes no assumptions about lead time and estimates overdiagnosis specific to life expectancy in addition to age alone.

There are some important caveats to the interpretation of these findings. First, our results were sensitive to the definition of screening mammography. We used a definition of screening mammography that may misclassify some diagnostic mammograms as screening. Using a more conservative definition—which may correctly categorize some diagnostic examinations and incorrectly categorize some screening examinations as nonscreening—estimates of overdiagnosis were smaller, ranging from 15% to 44% of cases. As a more conservative estimate, this approach offers a useful lower bound on risk for overdiagnosis. Even with this approach, we still saw that a substantial proportion of breast cancer cases among women with limited life expectancy or advanced age were overdiagnosed.

Second, the excess incidence estimated in this study incorporates the effect of screening patterns, including screening mammograms that occurred after cohort entry in each group, rather than from a single additional round of screening. Therefore, our estimates capture the risk for overdiagnosis associated with continued versus reduced screening, although with incomplete adherence (**Supplement Figure**). We would expect that overdiagnosis rates with perfect adherence to continuation or stopping screening might be even higher.

There are other important limitations to this work. This is an observational study and is subject to confounding. Women who choose to continue screening may be

at higher risk for developing breast cancer and lower risk for competing mortality. We adjusted for potential confounders, including age, race, and ethnicity, as well as factors that may influence competing risk. We also used cause-specific hazard models that may be useful for causal inference if there is differential competing risk for mortality (28). We could not adjust for breast density, family history, or other breast cancer risk factors as these are not observable in SEER-Medicare. Still, these factors may be of less importance in an older population where age is likely the most important risk factor and other traditional risk factors are less influential (37). We also specifically evaluated whether family history might play a substantial role in explaining our results and found that this is unlikely the main driver of our findings. Although we have used methods to address immortal time bias, we note that it is difficult to completely exclude this possibility. More broadly, methods used here tend to select for healthier patients both among screened and unscreened women. Our work uses an approach that requires lengthy follow-up to avoid labeling lead time as overdiagnosis. The results of our sensitivity analysis excluding screening mammograms performed more than 8 years after cohort entry (within 7 years of the end of follow-up) were similar to our main results. Furthermore, among women 85 years and older, most participants had died by the end of follow-up, making lead time an unlikely explanation for our findings. Lastly, we had limited power to evaluate benefits of screening, specifically potential reduction in breast cancer–specific mortality, and we did not evaluate other potential benefits of screening, such as reduction in invasive or burdensome treatments associated with earlier diagnosis.

In conclusion, women 70 years and older who continue breast cancer screening are at risk for overdiagnosis. The relative risk for overdiagnosis increases with age and is highest for the oldest women or those with lowest life expectancy. Overdiagnosis should be explicitly considered when making screening decisions, along with considering possible benefits of screening.

From Section of General Internal Medicine, Department of Internal Medicine, Yale School of Medicine; and Cancer Outcomes, Public Policy, and Effectiveness Research (COPPER) Center, Yale Cancer Center and Yale School of Medicine, New Haven, Connecticut (I.B.R., J.B.L., P.R.S., C.P.G.); and Cancer Outcomes, Public Policy, and Effectiveness Research (COPPER) Center, Yale Cancer Center and Yale School of Medicine; and Yale School of Public Health, New Haven, Connecticut (S.-Y.W.).

Note: The collection of cancer incidence data used in this study was supported by the California Department of Public Health pursuant to California Health and Safety Code Section 103885; Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries, under cooperative agreement 1NU58DP007156; the National Cancer Institute's SEER (Surveillance, Epidemiology, and End Results) program under contract HHSN261201800032I awarded to the University of California San Francisco, contract HHSN261201800015I awarded to the University of Southern California, and contract HHSN261201800009I awarded to the Public Health Institute.

Disclaimer: The ideas and opinions expressed herein are those of the author(s) and do not necessarily reflect the opinions of the State of California, Department of Public Health, the National Cancer Institute, and the Centers for Disease Control and Prevention or their contractors and subcontractors. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Acknowledgment: The authors gratefully acknowledge Meghan Lindsay, MPH, for her editorial assistance.

Grant Support: By the National Cancer Institute of the National Institutes of Health under award number K08CA248725 (Dr. Richman).

Disclosures: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M23-0133.

Reproducible Research Statement: *Study protocol:* Not available. *Statistical code:* Available on request to the corresponding author (ilana.richman@yale.edu). *Data set:* Not available due to privacy restrictions; however, access to SEER-Medicare may be requested independently.

Corresponding Author: Ilana B. Richman, MD, MHS, Yale School of Medicine, 367 Cedar Street, Harkness Hall A, Room 301a, New Haven, CT 06510; e-mail, ilana.richman@yale.edu.

Author contributions are available at Annals.org.

References

1. **National Center for Health Statistics.** Health, United States, 2016: With Chartbook on Long-term Trends in Health. National Center for Health Statistics; 2017. Report No. 2017-1232.
2. **Walter LC, Schonberg MA.** Screening mammography in older women: a review. JAMA. 2014;311:1336-1347. [PMID: 24691609] doi:10.1001/jama.2014.2834
3. **Nelson HD, Cantor A, Humphrey L, et al.** Screening for Breast Cancer: A Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Agency for Healthcare Research and Quality; 2016. Report No. 14-05201-EF-1.
4. **García-Albéniz X, Hernán MA, Logan RW, et al.** Continuation of annual screening mammography and breast cancer mortality in women older than 70 years. Ann Intern Med. 2020;172:381-389. [PMID: 32092767] doi:10.7326/M18-1199
5. **Lansdorp-Vogelaar I, Gulati R, Mariotto AB, et al.** Personalizing age of cancer screening cessation based on comorbid conditions: model estimates of harms and benefits. Ann Intern Med. 2014;161:104-112. [PMID: 25023249] doi:10.7326/M13-2867
6. **Schousboe JT, Sprague BL, Abraham L, et al.** Cost-effectiveness of screening mammography beyond age 75 years : a cost-effectiveness analysis. Ann Intern Med. 2022;175:11-19. [PMID: 34807717] doi:10.7326/M20-8076
7. **Siu AL; U.S. Preventive Services Task Force.** Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2016;164:279-296. [PMID: 26757170] doi:10.7326/M15-2886
8. **Qaseem A, Lin JS, Mustafa RA, et al; Clinical Guidelines Committee of the American College of Physicians.** Screening for breast cancer in average-risk women: a guidance statement from

the American College of Physicians. Ann Intern Med. 2019;170:547-560. [PMID: 30959525] doi:10.7326/M18-2147
9. **Oeffinger KC, Fontham ETH, Etzioni R, et al; American Cancer Society.** Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA. 2015;314:1599-1614. [PMID: 26501536] doi:10.1001/jama.2015.12783
10. **Welch HG, Fisher ES.** Diagnostic testing following screening mammography in the elderly. J Natl Cancer Inst. 1998;90:1389-1392. [PMID: 9747869] doi:10.1093/jnci/90.18.1389
11. **Braithwaite D, Zhu W, Hubbard RA, et al; Breast Cancer Surveillance Consortium.** Screening outcomes in older US women undergoing multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor characteristics or false positive rates? J Natl Cancer Inst. 2013;105:334-341. doi:10.1093/jnci/djs645 [23385442]
12. **Schonberg MA, Silliman RA, Marcantonio ER.** Weighing the benefits and burdens of mammography screening among women age 80 years or older. J Clin Oncol. 2009;27:1774-1780. [PMID: 19255318] doi:10.1200/JCO.2008.19.9877
13. **National Cancer Institute.** NCI Dictionary of Cancer Terms: overdiagnosis. Accessed at www.cancer.gov/publications/dictionaries/cancer-terms/def/overdiagnosis on 27 July 2022.
14. **Etzioni R, Gulati R.** Oversimplifying overdiagnosis. J Gen Intern Med. 2014;29:1218-1220. doi:10.1007/s11606-014-2867-0 [25948872]
15. **van Ravesteyn NT, Stout NK, Schechter CB, et al.** Benefits and harms of mammography screening after age 74 years: model estimates of overdiagnosis. J Natl Cancer Inst. 2015;107:djv103. doi:10.1093/jnci/djv103 [25948872]
16. **Zahl P-H, Jørgensen KJ, Gøtzsche PC, et al.** Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. Br J Cancer. 2013;109:2014-2019. [PMID: 23963144] doi:10.1038/bjc.2013.427
17. **Ripping TM, ten Haaf K, Verbeek ALM, et al.** Quantifying overdiagnosis in cancer screening: a systematic review to evaluate the methodology. J Natl Cancer Inst. 2017;109. [PMID: 29117353] doi:10.1093/jnci/djx060
18. **Welch HG, Prorok PC, O'Malley AJ, et al.** Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. N Engl J Med. 2016;375:1438-1447. [PMID: 27732805] doi:10.1056/NEJMoa1600249
19. **Independent UK Panel on Breast Cancer Screening.** The benefits and harms of breast cancer screening: an independent review. Lancet. 2012;380:1778-1786. [PMID: 23117178] doi:10.1016/S0140-6736(12)61611-0
20. **Killelea BK, Long JB, Chagpar AB, et al.** Evolution of breast cancer screening in the Medicare population: clinical and economic implications. J Natl Cancer Inst. 2014;106:dju159. [PMID: 25031307] doi:10.1093/jnci/dju159
21. **National Cancer Institute Division of Cancer Control & Population Sciences.** SEER-Medicare: Medicare Enrollment Data. Accessed at https://healthcaredelivery.cancer.gov/seermedicare/medicare/enroll.html on 5 March 2023.
22. **National Cancer Institute Division of Cancer Control & Population Sciences.** SEER-Medicare: Brief Description of the SEER-Medicare Database. Accessed at https://healthcaredelivery.cancer.gov/seermedicare/overview at on 5 March 2023.
23. **Hernán MA, Sauer BC, Hernández-Díaz S, et al.** Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016;79:70-75. [PMID: 27237061] doi:10.1016/j.jclinepi.2016.04.014
24. **Fenton JJ, Zhu W, Balch S, et al.** Distinguishing screening from diagnostic mammograms using Medicare claims data. Med Care. 2014;52:e44-51-e51. [PMID: 22922433] doi:10.1097/MLR.0b013e318269e0f5
25. **Kim DH, Schneeweiss S, Glynn RJ, et al.** Measuring frailty in Medicare data: development and validation of a claims-based frailty index. J Gerontol A Biol Sci Med Sci. 2018;73:980-987. [PMID: 29244057] doi:10.1093/gerona/glx229
26. **Tan A, Kuo Y-F, Goodwin JS.** Predicting life expectancy for community-dwelling older adults from Medicare claims data. Am J Epidemiol. 2013;178:974-983. [PMID: 23851579] doi:10.1093/aje/kwt054
27. **Fine JP, Gray RJ.** A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc. 1999;94:496-509.
28. **Rudolph JE, Lesko CR, Naimi AI.** Causal inference in the face of competing events. Curr Epidemiol Rep. 2020;7:125-131. [PMID: 33747714] doi:10.1007/s40471-020-00240-7
29. **Jackson LA, Jackson ML, Nelson JC, et al.** Evidence of bias in estimates of influenza vaccine effectiveness in seniors. Int J Epidemiol. 2006;35:337-344. [PMID: 16368725] doi:10.1093/ije/dyi274
30. **Lankham I, Slaughter MT.** A Framework for Simple and Efficient Bootstrap Validation in SAS®, with Examples. PharmaSUG; 2022. Paper SA-149. Accessed at www.pharmasug.org/proceedings/2022/SA/PharmaSUG-2022-SA-149.pdf on 1 May 2023.
31. **Ryser MD, Lange J, Inoue LYT, et al.** Estimation of breast cancer overdiagnosis in a U.S. breast screening cohort. Ann Intern Med. 2022;175:471-478. [PMID: 35226520] doi:10.7326/M21-3577
32. **Taylor LJ, Steiman JS, Anderson B, et al.** Does persistent use of radiation in women > 70 years of age with early-stage breast cancer reflect tailored patient-centered care? Breast Cancer Res Treat. 2020;180:801-807. [PMID: 32125557] doi:10.1007/s10549-020-05579-5
33. **Frebault J, Bergom C, Cortina CS, et al.** Invasive breast cancer treatment patterns in women age 80 and over: a report from the National Cancer Database. Clin Breast Cancer. 2022;22:49-59. [PMID: 34391660] doi:10.1016/j.clbc.2021.07.005
34. **Minami CA, Jin G, Freedman RA, et al.** Association of surgery with frailty status in older women with early-stage breast cancer. JAMA Surg. 2023;158:664-666. [PMID: 36920410] doi:10.1001/jamasurg.2022.8146
35. **Yoo GJ, Levine EG, Aviv C, et al.** Older women, breast cancer, and social support. Support Care Cancer. 2010;18:1521-1530. [PMID: 20967554] doi:10.1007/s00520-009-0774-4
36. **Lannin DR, Wang S.** Are small breast cancers good because they are small or small because they are good? N Engl J Med. 2017;376:2286-2291. [PMID: 28591529] doi:10.1056/NEJMsr1613680
37. **Vacek PM, Skelly JM, Geller BM.** Breast cancer risk assessment in women aged 70 and older. Breast Cancer Res Treat. 2011;130:291-299. [PMID: 21604157] doi:10.1007/s10549-011-1576-1

Author Contributions: Conception and design: I.B. Richman, C. P. Gross.
Analysis and interpretation of the data: I.B. Richman, J.B. Long, P.R. Soulos, S.-Y. Wang, C.P. Gross.
Drafting of the article: I.B. Richman.
Critical revision of the article for important intellectual content: I.B. Richman, J.B. Long, P.R. Soulos, S.-Y. Wang, C.P. Gross.
Final approval of the article: I.B. Richman, J.B. Long, P.R. Soulos, S.-Y. Wang, C.P. Gross.
Statistical expertise: I.B. Richman, J.B. Long, P.R. Soulos.
Obtaining of funding: I.B. Richman.
Administrative, technical, or logistic support: I.B. Richman, C.P. Gross.
Collection and assembly of data: J.B. Long, P.R. Soulos.