Estimating Breast Cancer Overdiagnosis after Screening Mammography among Older Women in the US

Ilana B. Richman, MD MHS[1,2], **Jessica B. Long, MPH**[1,2], **Pamela R. Soulos, MPH**[1,2], **Shi-Yi Wang, MD PhD**[2,3], **Cary P. Gross, MD**[1,2]

[1.] Department of Internal Medicine, Section of General Internal Medicine, Yale School of Medicine, New Haven, CT

[2.] Cancer Outcomes, Public Policy, and Effectiveness Research (COPPER) Center, Yale Cancer Center and Yale University School of Medicine, New Haven, CT

[3.] Yale School of Public Health, New Haven, CT

Abstract

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

Objective: To estimate risk of 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 to those who did not. Analyses used competing risk models, stratified by age.

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

Patients: Women 70 and older with no history of breast cancer and who had been recently screened.

Exposure: Continued screening in the next interval.

Measurements: Breast cancer diagnoses and breast cancer death over 15 years follow up.

Corresponding Author Ilana Richman, 367 Cedar St, New Haven CT, 06510, Ilana.richman@yale.edu.

Results: This study included 54,635 women. Among women aged 70–74, the adjusted cumulative incidence of breast cancer was 6.1 cases per 100 screened women (95% CI 5.7–6.4) versus 4.2 cases per 100 unscreened women (95% CI 3.4–5.2). An estimated 31% of breast cancers among screened women were potentially overdiagnosed. For women ages 75–84, cumulative incidence was 4.9 per 100 screened women (95% CI 4.6–5.1) versus 2.6 per 100 unscreened women (95% CI 2.2–3.0) among those not, with 47% of cases potentially overdiagnosed. For women ages 85 and older, the cumulative incidence was 2.8 (95% CI 2.3–3.4) among screened women versus 1.5 (95% CI 0.6–2.2), with up to 54% overdiagnosis. We did not observe 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 of 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 many older women who are diagnosed with breast cancer after screening may be overdiagnosed, especially among the oldest women. Whether harms of overdiagnosis are balanced by benefits and for whom remains an important question.

Introduction:

Although older women are commonly screened for breast cancer, the efficacy of screening in this population remain uncertain (1). No randomized trials have evaluated screening mammography in women 75 and older, and only a few studies have included women over the age of 70, leaving uncertainty about benefits and harms of screening in this group (2,3). Observational studies suggest that the mortality benefit from screening may be limited to women younger than 75.(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 US Preventive Services Task Force makes no specific recommendation for or against screening women 75 and older, but includes women 70–74 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, while the American College of Physicians recommends discontinuing screening at age 75 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 of overdiagnosis is driven by several factors, including the biological behavior of a tumor and life expectancy (14). Specifically, some breast cancers may have a long pre-symptomatic phase. Detecting these breast cancers through screening may result in overdiagnosis if these breast cancers

would have otherwise remained clinically silent during a patient's lifetime. Additionally, even aggressive breast cancers with a short pre-symptomatic phase may be overdiagnosed in older women who have very limited life expectancy. Modeling studies have estimated that overdiagnosis may occur in approximately 0.8–7.5 per 1000 older women screened for breast cancer depending on age, and may account for between 12–48% of screen-detected breast cancers (5,15). Although important, modeling studies have some inherent limitations. For example, modeling studies make assumptions about the distribution of lead times of breast cancers, which are not directly observable (16,17). Studies using alternate methods including long term follow up of randomized trials have often focused younger screened populations rather than older women (18,19).

The primary goal of this study was to quantify the risk of 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 among women 70 and older. To do this, we approximated a target trial in which women 70 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 one more round or not at the time of their next mammogram. Anticipating heterogeneity among women 70 and older, we stratified by age (70–74, 75–84, and 85 and older). Since 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 Surveillance, Epidemiology, and End Results (SEER)-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 cancers, 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 70 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 age 70 or older by January 1, 2003, had not had breast cancer prior to 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. 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 non-screening arm if they received a non-screening mammogram, were diagnosed with breast cancer, or died before the pseudomammogram date, but not after. Therefore, women in both arms (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 et al. which distinguishes screening mammograms from diagnostic mammograms in claims data (eMethods, eTable 1) (24). The algorithm has a sensitivity of 99.7% and a specificity of 69.4% 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 three years following their 2002 mammogram. Women who did not have a screening mammogram during this timeframe were included in the non-screening 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 arm. Secondary outcomes included breast cancer diagnosis by stage (in-situ, invasive localized and regional/distant cancers) 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 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 a previously 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 of 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 (70–74, 75–84, 85+) or life expectancy at cohort entry (5 years, 6–10 years, >10 years). We adjusted models for variables which may influence both screening use and the underlying risk of breast cancer, specifically age, race, and ethnicity. We also adjusted models for factors which may influence both screening use and competing risk of 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 1,000 replicate samples to estimate 95% confidence intervals for our estimates (30). Lastly, we quantified the risk of overdiagnosis among screened women. 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 mammograms correctly as diagnostic or screening, 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 (eMethods). 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 the Fine-Gray model. 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 > 8 years after cohort entry in order to ensure sufficient

follow up time to observe breast cancer diagnoses(31). Lastly, we evaluated the potential impact of family history as an unmeasured confounder on our results (eMethods).

Results:

The cohort included 54,635 women (Table 1). The mean age of the population was 77.2 years (95% confidence interval 77.1–77.2), 6% of women were Black, 3% were Hispanic, and 88% were White. Life expectancy was 10 years 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 be considered frail (Table 1).

Among women ages 70–74 years, 88% were screened at cohort entry (i.e., within 3 years of the 2002 mammogram). Among women ages 75–84, 81% were screened at cohort entry, and among women ages 85 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 70–74 who were not screened at cohort entry, 30% were screened in the first 3 years of follow up (eFigure 2a). Among women with a life expectancy of 75–84 years not screened at cohort entry, 16% were screened in the first 3 years (eFigure 2b). For women ages 85+, 6% were screened in the first 3 years of follow up (eFigure 2c).

Among women ages 70–74, median follow up time was 13.7 years (IQR 9.2–14.4), 10 (5.8–13.9) years for women ages 75–84 and 5.7 (3.1–9.1) years for women 85+. By the end of follow up, among those 70–74 who were screened 38% had died, versus 56% among those who were not screened. Among those ages 75–84, 65% of those who were screened had died vs 80% among those who were not screened. For those ages 85 and older, 91% of those screened had died vs 96% among those not screened.

In adjusted analyses using Fine-Gray competing risk models, the cumulative incidence of breast cancer was 6.1 cases per 100 women (95% CI 5.7–6.4) among those 70–74 who were screened at cohort entry, versus 4.2 cases per 100 women (95% CI 3.4–5.0) among those who were not screened at cohort entry (risk difference 1.9 cases per 100 (95% CI 1.0–2.8)) (Figure 2, Table 2). Among women screened at cohort entry who were eventually diagnosed with breast cancer, we estimated 31% may be overdiagnosed. Among women ages 75–84, the cumulative incidence of breast cancer was 4.9 per 100 (95% CI 4.6–5.2) among women who were screened at cohort entry versus 2.6 per 100 (95% CI 2.2–3.0) among women who were not screened at cohort entry (risk difference 2.3 (95% CI 1.7–2.8), Figure 2, Table 2). We estimated that 47% of breast cancer cases among screened women may be overdiagnosed. For women 85 and older who were screened, cumulative incidence of breast cancer was 2.8 per 100 (95% CI 2.3–3.4) versus 1.3 per 100 (95% CI 0.9–1.8) among women not screened at cohort entry (risk difference 1.5 (95% CI 0.6–2.2) Figure 2, Table 2). Risk of overdiagnoses was estimated at 54% among screened women diagnosed with breast cancer. When stratifying by life expectancy, an estimated 32% of breast cancers among screened women with a life expectancy of >10 years were overdiagnosed. Among women with a life expectancy of 6–10 years, 53% of cancers were potentially overdiagnosed, and 62% among women with a life expectancy 5 years (eTable 3).

In sensitivity analyses, models using logistic regression, and models that censored women if screening was performed > 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 70–74 was 0.9 breast cancer cases per 100 (95% CI –0.1–1.7), with an estimated 15% of screen-detected cancers overdiagnosed. For women ages 75–84, the risk difference was 1.7 (95% CI 1.1–2.2) with an estimated 36% of breast cancer cases overdiagnosed. For women ages 85 and older, the risk difference was 1.1 (95% CI 0.3–1.7) after 15 years of follow up with an estimated 44% of screen detected cancers 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 (eTable 4). Estimates of the impact of family history suggested differential screening use among women with a first degree relative with breast cancer would not explain our results (eTable 5).

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

Discussion:

We found that the proportion of breast cancers that may be overdiagnosed among older women who are screened is considerable, and rises with advancing age and with decreasing life expectancy. For women 85 and older, 54% of breast cancers among screened women may be overdiagnosed. For younger women, 70–74, the proportion is smaller but still considerable at up to 31%. We also observed that the absolute risk of overdiagnosis was similar across strata of life expectancy and ranged from 1.5–2.3 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 of overdiagnosis of about 2% over 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 observe 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, 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 (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 cancers rather than advanced breast cancers. Whether overdiagnosis of these early-stage cancers is consequential in part depends on whether diagnosis results in aggressive or burdensome treatments. Up to 90% of women aged 80 and older with non-metastatic breast cancers undergo surgery and nearly two thirds of women over 70 have radiation for early stage invasive breast cancers (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 screening is associated with lower rates of chemotherapy use, which is an important potential benefit of screening that must be weighed against the risks of 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 lead time for breast cancer is about 7 years. Given this, more than half of breast cancers identified among women with a mean life expectancy <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 may be overdiagnosed and 54% of cases among women age 85 or more years may be overdiagnosed. Our main results were somewhat higher than a modeling study which estimated that between 12–48% of screen-detected cancers among women 75 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 cancers among women over the age of 80 are 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 exams and incorrectly categorize some screening exams as non-screening --estimates of overdiagnosis were smaller, ranging from 15–44% of cases. As a more conservative estimate, this approach offers a useful lower bound on risk of overdiagnosis. Even with this approach, we still observed 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 observed, including screening mammograms that occurred after cohort entry in each arm, rather than from a single additional round of screening. Therefore, our estimates

capture the risk of overdiagnosis associated with continued versus reduced screening, although with incomplete adherence (eFig2). 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 of developing breast cancer and lower risk of 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 which may be useful for causal inference if there is differential competing risk of 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 single most potent risk factor (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 > 8 years after cohort entry (within 7 years of the end of follow up) were similar to our main results. Further, among women 85 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 a as reduction in invasive or burdensome treatments associated with earlier diagnosis.

Conclusions:

Women 70 and older who continue breast cancer screening are at risk of overdiagnosis. The relative risk of 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

References:

1. No Title. In: Health, United States, 2016: With Chartbook on Long-term Trends in Health [Internet]. Hyattsville (MD); 2017. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28910066

2. Walter LC, Schonberg MA. Screening mammography in older women: a review. JAMA [Internet]. 2014;311(13):1336–47. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24691609 [PubMed: 24691609]

3. Nelson HD, Cantor A, Humphrey L, Fu R, Pappas M, Daeges M, et al. Screening for Breast Cancer [Internet]. Screening for Breast Cancer: A Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Agency for Healthcare Research and Quality (US); 2016 [cited 2017 Feb 22]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26889531

4. García-Albéniz X, Hernán MA, Logan RW, Price M, Armstrong K, Hsu J. Continuation of Annual Screening Mammography and Breast Cancer Mortality in Women Older Than 70 Years. Ann Intern Med [Internet]. 2020; Available from: 10.7326/M18-1199

5. Lansdorp-Vogelaar I, Gulati R, Mariotto AB, Schechter CB, de Carvalho TM, Knudsen AB, et al. Personalizing age of cancer screening cessation based on comorbid conditions: model estimates of harms and benefits. Ann Intern Med [Internet]. 2014;161(2):104–12. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25023249 [PubMed: 25023249]

6. Schousboe JT, Sprague BL, Abraham L, O'Meara ES, Onega T, Advani S, et al. Cost-Effectiveness of Screening Mammography Beyond Age 75 Years : A Cost-Effectiveness Analysis. Ann Intern Med. 2022 Jan;175(1):11–9. [PubMed: 34807717]

7. Siu AL. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med [Internet]. 2016 Jan 12 [cited 2016 Jan 13];164(4):279–96. Available from: http://annals.org/article.aspx?articleid=2480757 [PubMed: 26757170]

8. Qaseem A, Lin JS, Mustafa RA, Horwitch CA, Wilt TJ. Screening for Breast Cancer in Average-Risk Women: A Guidance Statement From the American College of Physicians. Ann Intern Med [Internet]. 2019 Apr 9;170(8):547–60. Available from: 10.7326/M18-2147 [PubMed: 30959525]

9. Oeffinger KC, Fontham ETH, Etzioni R, Herzig A, Michaelson JS, Shih Y-CT, et al. Breast Cancer Screening for Women at Average Risk. JAMA [Internet]. 2015 Oct 20 [cited 2017 May 30];314(15):1599. Available from: 10.1001/jama.2015.12783 [PubMed: 26501536]

10. Welch HG, Fisher ES. Diagnostic testing following screening mammography in the elderly. J Natl Cancer Inst [Internet]. 1998;90(18):1389–92. Available from: https://www.ncbi.nlm.nih.gov/pubmed/9747869 [PubMed: 9747869]

11. Schonberg MA, McCarthy EP. Mammography screening among women age 80 years and older: consider the risks. J Clin Oncol [Internet]. 2009;27(4):640–2. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19075269 [PubMed: 19075269]

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 [Internet]. 2009;27(11):1774–80. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19255318 [PubMed: 19255318]

13. National Cancer Institute [Internet]. [cited 2022 Jul 27]. Available from: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/overdiagnosis

14. Etzioni R, Gulati R. Oversimplifying overdiagnosis. J Gen Intern Med [Internet]. 2014;29(9):1218–20. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24830739 [PubMed: 24830739]

15. van Ravesteyn NT, Stout NK, Schechter CB, Heijnsdijk EAM, Alagoz O, Trentham-Dietz A, et al. Benefits and Harms of Mammography Screening After Age 74 Years: Model Estimates of Overdiagnosis. JNCI J Natl Cancer Inst [Internet]. 2015 Jul 1;107(7):djv103. Available from: 10.1093/jnci/djv103 [PubMed: 25948872]

16. Zahl P-H, Jørgensen KJ, Gøtzsche P. Zahl PH, Jørgensen KJ, Gøtzsche PCOverestimated lead times in cancer screening has lead substantial underestimation of overdiagnosis. Br J Cancer 109: 2014–2019. Br J Cancer. 2013 Aug 20;109.

17. Ripping TM, ten Haaf K, Verbeek ALM, van Ravesteyn NT, Broeders MJM. Quantifying Overdiagnosis in Cancer Screening: A Systematic Review to Evaluate the Methodology. JNCI J Natl Cancer Inst [Internet]. 2017 Oct 1;109(10):djx060. Available from: 10.1093/jnci/djx060

18. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. N Engl J Med [Internet]. 2016;375(15):1438–47. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27732805 [PubMed: 27732805]

19. The benefits and harms of breast cancer screening: an independent review. Lancet (London, England). 2012 Nov;380(9855):1778–86. [PubMed: 23117178]

20. Killelea BK, Long JB, Chagpar AB, Ma X, Wang R, Ross JS, et al. Evolution of breast cancer screening in the Medicare population: clinical and economic implications. J Natl Cancer Inst [Internet]. 2014;106(8). Available from: https://www.ncbi.nlm.nih.gov/pubmed/25031307

21. Medicare Enrollment Data - SEER-Medicare [Internet]. [cited 2023 Mar 5]. Available from: https://healthcaredelivery.cancer.gov/seermedicare/medicare/enroll.html

22. National Cancer Institute. SEER-Medicare: Brief Description of the SEER-Medicare Database [Internet]. Available from: https://healthcaredelivery.cancer.gov/seermedicare/overview/

23. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol [Internet]. 2016;79:70–5. Available from: https://www.sciencedirect.com/science/article/pii/S0895435616301366 [PubMed: 27237061]

24. Fenton JJ, Zhu W, Balch S, Smith-Bindman R, Fishman P, Hubbard RA. Distinguishing screening from diagnostic mammograms using Medicare claims data. Med Care [Internet]. 2014;52(7):e44–51. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22922433 [PubMed: 22922433]

25. Kim DH, Schneeweiss S, Glynn RJ, Lipsitz LA, Rockwood K, Avorn J. Measuring Frailty in Medicare Data: Development and Validation of a Claims-Based Frailty Index. J Gerontol A Biol Sci Med Sci. 2018 Jun;73(7):980–7. [PubMed: 29244057]

26. Tan A, Kuo YF, Goodwin JS. Predicting life expectancy for community-dwelling older adults from Medicare claims data. Am J Epidemiol [Internet]. 2013;178(6):974–83. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23851579 [PubMed: 23851579]

27. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J Am Stat Assoc [Internet]. 1999 Jun 1;94(446):496–509. Available from: 10.1080/01621459.1999.10474144

28. Rudolph JE, Lesko CR, Naimi AI. Causal inference in the face of competing events. Curr Epidemiol reports. 2020 Sep;7(3):125–31.

29. Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. Int J Epidemiol. 2006 Apr;35(2):337–44. [PubMed: 16368725]

30. A Framework for Simple and Efficient Bootstrap Validation in SAS®, with Examples [Internet]. [cited 2023 May 1]. Available from: https://www.pharmasug.org/proceedings/2022/SA/PharmaSUG-2022-SA-149.pdf

31. Ryser MD, Lange J, Inoue LYT, O'Meara ES, Gard C, Miglioretti DL, et al. Estimation of Breast Cancer Overdiagnosis in a U.S. Breast Screening Cohort. Ann Intern Med [Internet]. 2022 Mar 1;175(4):471–8. Available from: 10.7326/M21-3577 [PubMed: 35226520]

32. Taylor LJ, Steiman JS, Anderson B, Schumacher JR, Wilke LG, Greenberg CC, 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 Apr;180(3):801–7. [PubMed: 32125557]

33. Frebault J, Bergom C, Cortina CS, Shukla ME, Zhang Y, Huang C-C, et al. Invasive Breast Cancer Treatment Patterns in Women Age 80 and Over: A Report from the National Cancer Database. Clin Breast Cancer 2022 Jan;22(1):49–59. [PubMed: 34391660]

34. Minami CA, Jin G, Freedman RA, Schonberg MA, King TA, Mittendorf EA. Association of Surgery With Frailty Status in Older Women With Early-Stage Breast Cancer. JAMA Surg [Internet]. 2023 Mar 15; Available from: 10.1001/jamasurg.2022.8146

35. Yoo GJ, Levine EG, Aviv C, Ewing C, Au A. Older women, breast cancer, and social support. Support care cancer Off J Multinatl Assoc Support Care Cancer. 2010 Dec;18(12):1521–30.

36. Lannin DR, Wang S. Are Small Breast Cancers Good because They Are Small or Small because They Are Good? N Engl J Med [Internet]. 2017;376(23):2286–91. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28591529 [PubMed: 28591529]

37. Vacek PM, Skelly JM, Geller BM. Breast cancer risk assessment in women aged 70 and older. Breast Cancer Res Treat. 2011 Nov;130(1):291–9. [PubMed: 21604157]
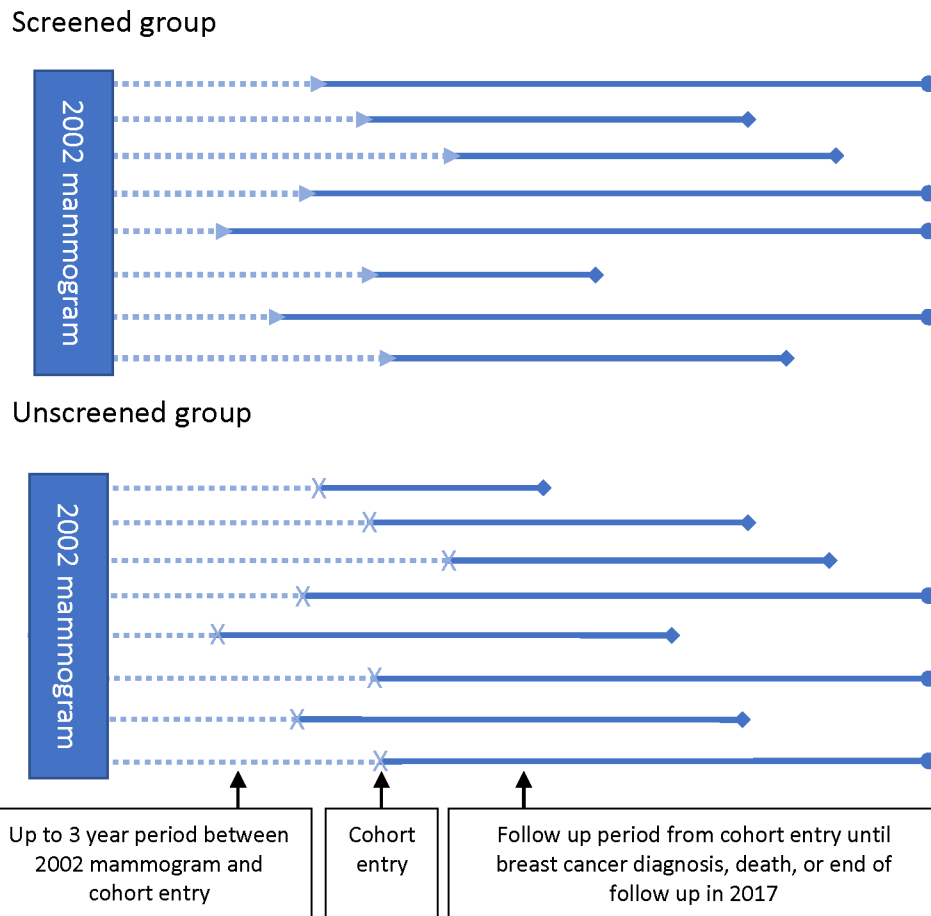
Figure 1: Study schematic.

Each horizontal line represents an individual in the study. Study entry begins at the date of the mammogram (triangular arrowhead) or pseudomammogram (x shape), which must be within 3 years of the 2002 mammogram. The time between the 2002 mammogram and study entry are similar for both the screened and unscreened group, 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 (diamond) or end of follow up in 2017 (circle).

Panel A: Cumulative incidence of breast cancer, ages 70-74 years
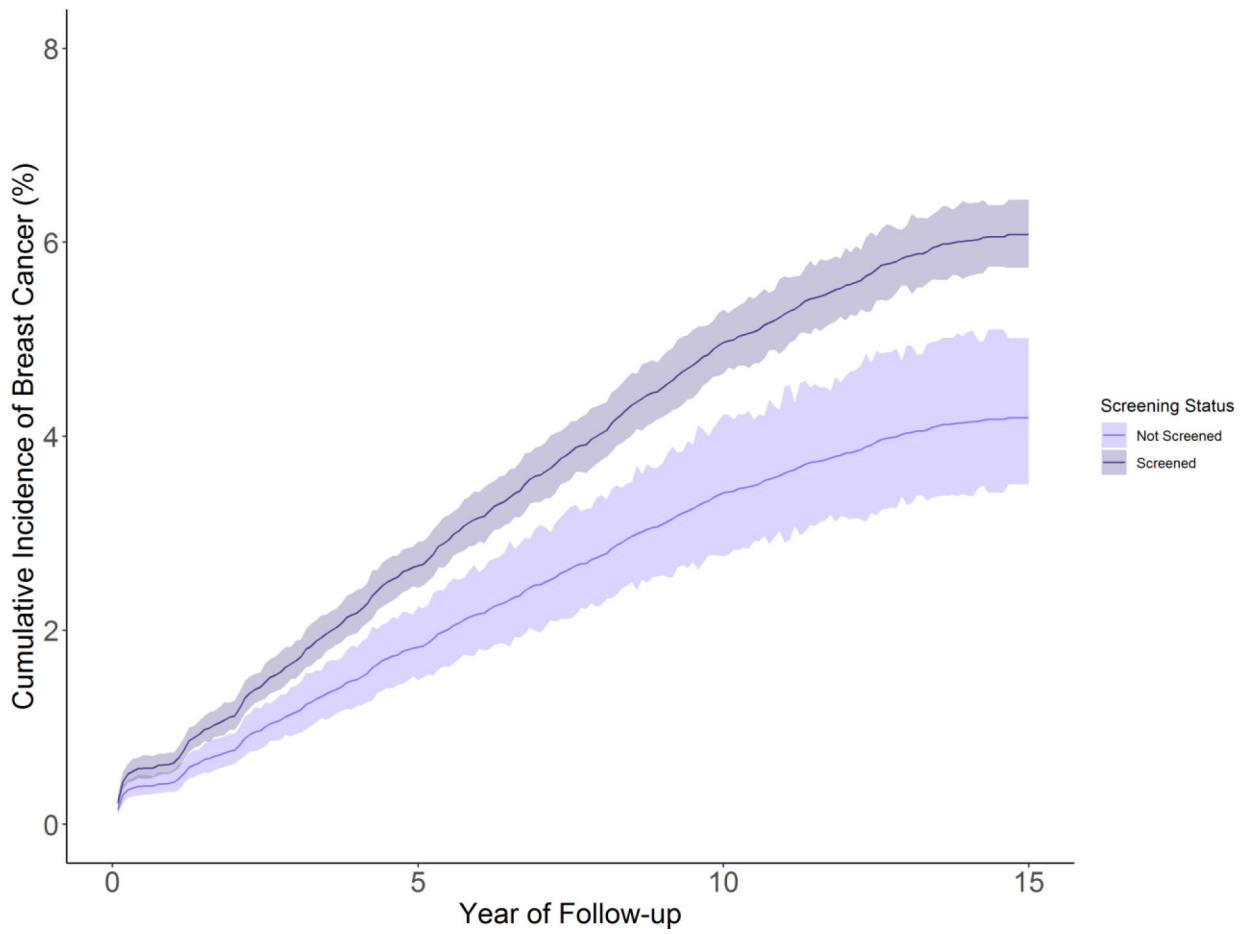
Panel B: Cumulative incidence of breast cancer, ages 75-84 years
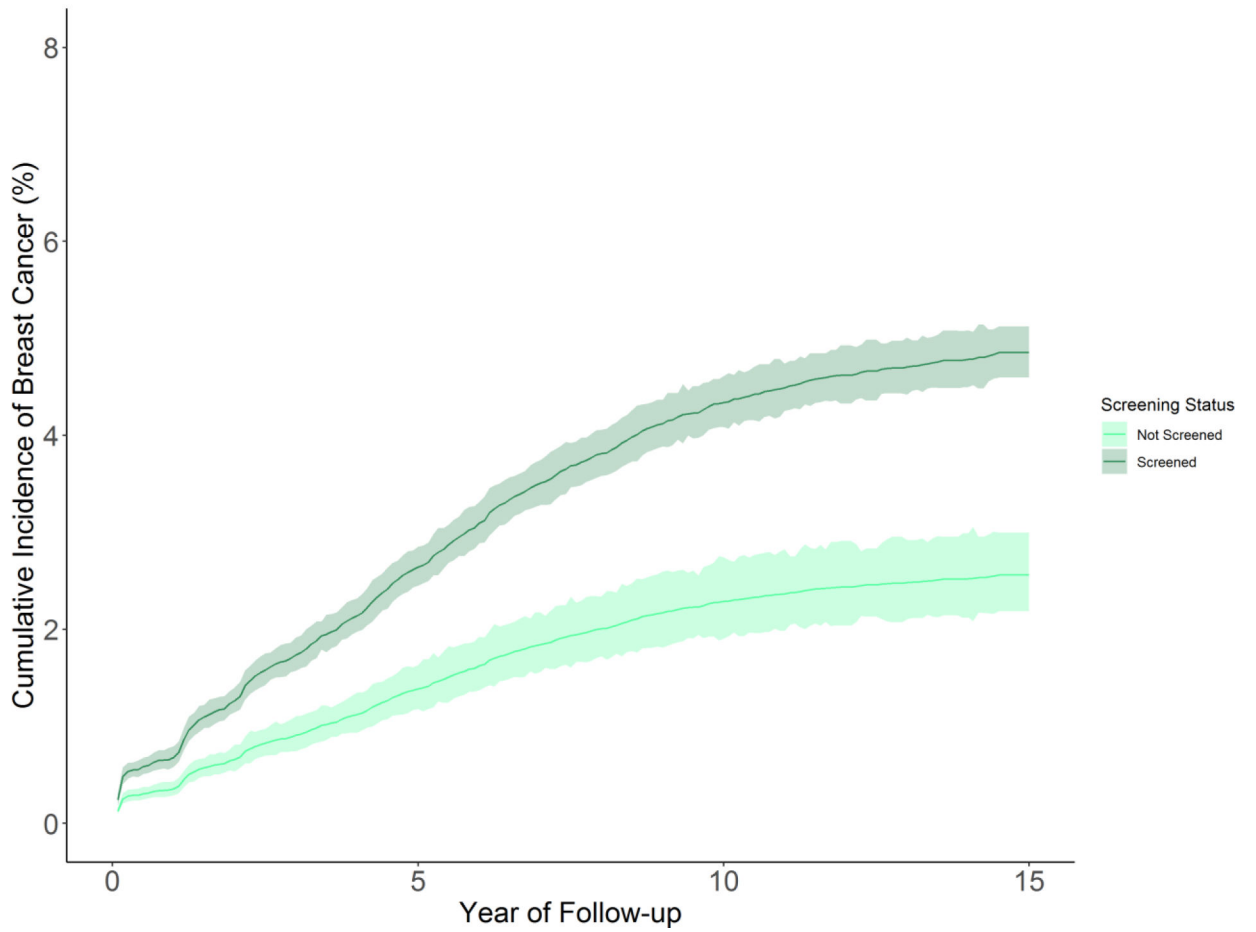
Panel C: Cumulative incidence of breast cancer, ages 85+ years
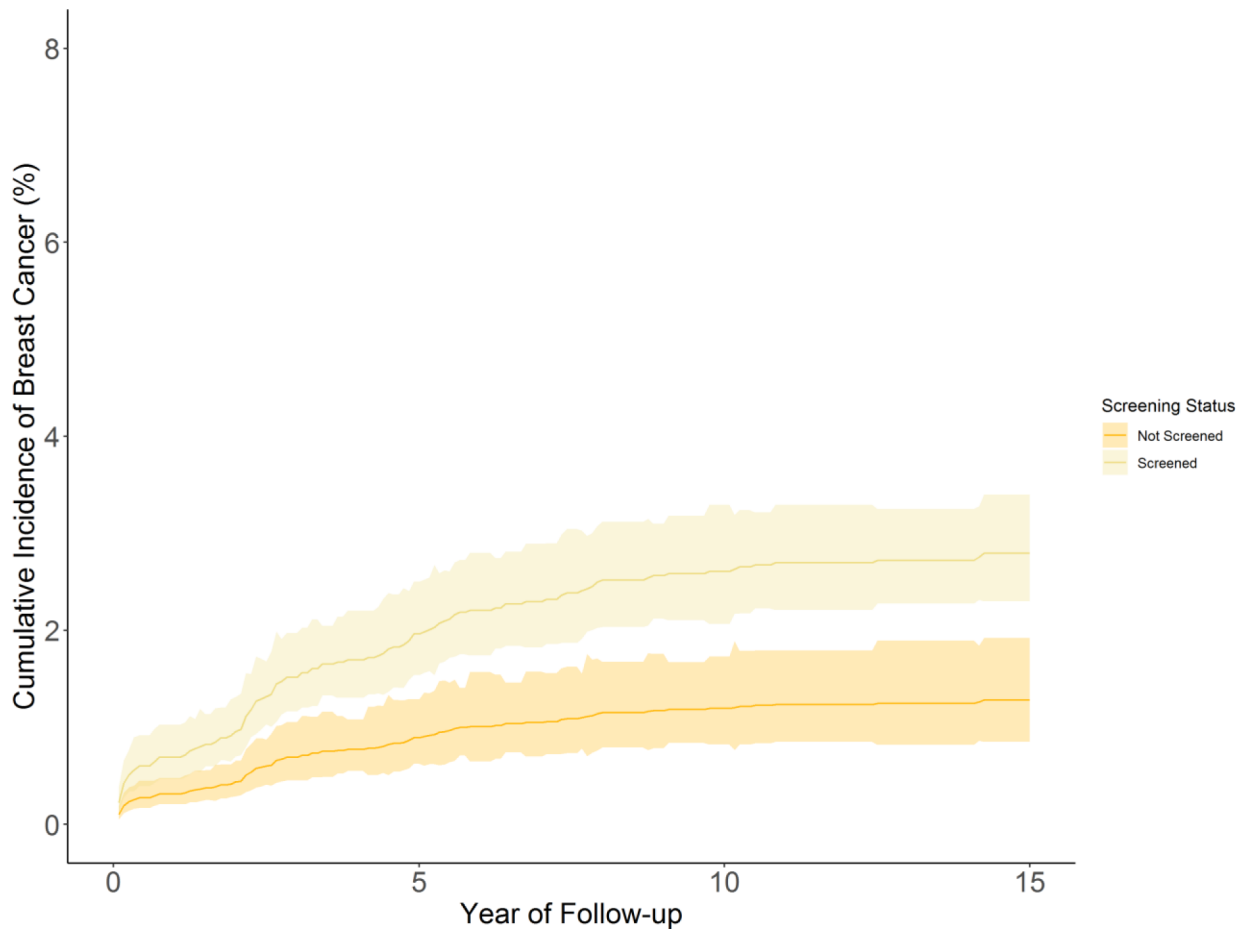
Figure 2: Cumulative incidence of breast cancer by screening status and age.
Figure panels depict cumulative incidence of breast cancer (breast cancer cases per 100 women) among women screened or not screened at cohort entry over available follow up. Shaded areas indicate 95% confidence intervals. Panel A: Age 70–74, Panel B: Age 75–84 years, Panel C: Age 85 and older.

Table 1:

Descriptive characteristics of study sample

|	Age 70–74			Age 75–84			Age 85+		
	Screened	Unscreened	Standardized Difference	Screened	Unscreened	Standardized Difference	Screened	Unscreened	Standardized Difference
Number of Beneficiaries	17,488	2,437		23,613	5,707		3,384	2,006	
Age, mean (95% CI)	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**
Life Expectancy in months, mean (95% CI)	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 Years	116 (1)	62 (3)	**0.15**	1263 (5)	666 (12)	**0.23**	1418 (42)	1069 (53)	**0.23**
LE 6–10 Years	1055 (6)	280 (11)	**0.19**	10484 (44)	2976 (52)	**0.16**	1966 (58)	937 (47)	**0.23**
LE > 10 Years	16317 (93)	2095 (86)	**0.24**	11866 (50)	2065 (36)	**0.29**	NA	NA	NA
Race and Ethnicity									
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	15184 (87)	1936 (79)	**0.20**	21258 (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									
0	10533 (60)	1186 (49)	**0.23**	12723 (54)	2494 (44)	**0.20**	1547 (46)	819 (41)	**0.10**
1 to 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 months	10322 (59)	1152 (47)	**0.24**	14791 (63)	3132 (55)	**0.16**	2190 (65)	1176 (59)	**0.13**
Primary care provider visit in prior 12 months	14451 (83)	1944 (80)	0.07	19716 (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									
< 5%	4073 (23)	450 (18)	**0.12**	5597 (24)	1207 (21)	0.06	734 (22)	444 (22)	0.01
5–9.9%	5433 (31)	703 (29)	0.05	7496 (32)	1678 (29)	0.05	1106 (33)	586 (29)	0.08

	Age 70–74			Age 75–84			Age 85+		
	Screened	Unscreened	Standardized Difference	Screened	Unscreened	Standardized Difference	Screened	Unscreened	Standardized Difference
10–19.9%	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	1538 (9)	460 (19)	**0.30**	1761 (7)	904 (16)	**0.26**	269 (8)	245 (12)	**0.14**
Nonmetro residence	2942 (17)	441 (18)	0.03	3559 (15)	942 (17)	0.04	473 (14)	289 (14)	0.01

Notes: Column values are N (column percent) except for age and life expectancy, where values indicate mean (95% confidence interval).

Bolded values 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. 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 algorithm by Kim, et al., 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. NA=not applicable

Table 2:

Cumulative incidence of breast cancer cases per 100 individuals

Age	Exposure	Primary Analysis		Sensitivity Analyses			
		Unadjusted	Adjusted	Cause specific	Logistic Regression	Censored if screened > 8 years after cohort entry	Alternate Screening Definition
70–74	Not Screened	4.0 (3.3–4.9)	4.2 (3.5–5.0)	5.5 (4.6–6.6)	4.2 (3.4–5.1)	3.6 (2.9–4.5)	4.9 (4.2–5.8)
	Screened	6.2 (5.9–6.6)	6.1 (5.7–6.4)	7.1 (6.6–7.5)	6.0 (5.7–6.4)	5.1 (4.8–5.5)	5.8 (5.5–6.2)
	Difference	2.2 (1.3–3.0)	1.9 (1.0–2.8)	1.6 (0.4–2.7)	1.9 (1.0–2.7)	1.5 (0.6–2.3)	0.9 (−0.1–1.7)
	% Excess	35	31	22	31	29	15
	Hazard Ratio	1.56 (1.27–1.91)	1.47 (1.19–1.81)	1.29 (1.05–1.59)	1.48 (1.2–1.83)	1.41 (1.12–1.78)	1.19 (0.99–1.43)
75–84	Not Screened	2.4 (2.1–2.8)	2.6 (2.2–3.0)	4.1 (3.4–4.8)	2.6 (2.2–3.0)	2.3 (1.9–2.7)	3 (2.7–3.5)
	Screened	5.0 (4.8–5.3)	4.9 (4.6–5.2)	6.4 (6.0–6.8)	4.8 (4.5–5.1)	4.4 (4.1–4.6)	4.7 (4.4–5)
	Difference	2.6 (2.1–3.1)	2.3 (1.7–2.8)	2.3 (1.5–3.1)	2.3 (1.7–2.8)	2.1 (1.6–2.6)	1.7 (1.1–2.2)
	% Excess	52	47	36	47	47	36
	Hazard Ratio	2.10 (1.76–2.50)	1.92 (1.6–2.3)	1.59 (1.33–1.91)	1.93 (1.61–2.31)	1.93 (1.59–2.33)	1.56 (1.32–1.83)
85+	Not Screened	1.3 (0.9–2)	1.3 (0.9–1.9)	3.2 (2.0–5.1)	1.3 (0.9–1.9)	1.3 (0.9–1.9)	1.4 (1.0–2.0)
	Screened	2.9 (2.3–3.7)	2.8 (2.3–3.4)	5.6 (4.2–7.5)	2.8 (2.3–3.4)	2.7 (2.2–3.4)	2.5 (1.9–3.1)
	Difference	1.6 (0.8–2.4)	1.5 (0.6–2.2)	2.4 (0.6–4.2)	1.5 (0.6–2.2)	1.4 (0.6–2.0)	1.1 (0.3–1.7)
	% Excess	55	54	43	53	52	44
	Hazard Ratio	2.56 (1.46–3.47)	2.2 (1.43–3.4)	1.78 (1.15–2.76)	2.15 (1.39–3.33)	2.13 (1.37–3.29)	1.76 (1.15–2.69)

Notes: Tables present the cumulative incidence of breast cancer (breast cancer cases per 100 individuals) at the end of follow up, which occurred at death, breast cancer diagnosis, or through the end of 2017. All values in parenthesis indicate 95% confidence intervals. Hazard ratios compare risk of breast cancer diagnosis in screened and unscreened groups. Logistic models produce odds ratios rather than hazard ratios. All models use the Fine-Gray method, except the logistic model and the cause-specific hazard model. 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."

Table 3:

Adjusted cumulative incidence of stage-specific breast cancer and breast cancer death per 100 individuals

Age	Exposure	Overall Breast Cancer Incidence	In Situ Breast Cancer Incidence	Localized Invasive Breast Cancer Incidence	Regional-Distant Breast Cancer Incidence	Breast Cancer Mortality
70–74	Not Screened	4.19 (3.49–5.03)	0.59 (0.36–0.98)	2.56 (2.05–3.20)	0.90 (0.61–1.34)	0.41 (0.22–0.76)
	Screened	6.08 (5.74–6.44)	1.09 (0.94–1.27)	3.84 (3.58–4.11)	1.00 (0.85–1.17)	0.35 (0.26–0.48)
	Difference	1.89 (0.98–2.75)	0.50 (0.10–0.81)	1.28 (0.51–1.93)	0.10 (−0.31–0.47)	−0.06 (−0.34–0.16)
	% Excess	31	46	33	10	−17
	Hazard Ratio	1.47 (1.19–1.81)	1.86 (1.08–3.18)	1.51 (1.15–1.98)	1.11 (0.71–1.72)	0.86 (0.44–1.68)
75–84	Not Screened	2.56 (2.20–2.97)	0.15 (0.07–0.29)	1.50 (1.21–1.86)	0.74 (0.55–1)	0.42 (0.28–0.64)
	Screened	4.85 (4.57–5.15)	0.79 (0.68–0.93)	3.15 (2.95–3.38)	0.78 (0.66–0.92)	0.36 (0.29–0.46)
	Difference	2.29 (1.74–2.81)	0.64 (0.46–0.79)	1.65 (1.21–2.03)	0.04 (−0.21–0.27)	−0.06 (−0.27–0.11)
	% Excess	47	81	52	5	−17
	Hazard Ratio	1.92 (1.60–2.3)	5.41 (2.65–11.06)	2.12 (1.68–2.67)	1.05 (0.75–1.49)	0.87 (0.55–1.37)
85+	Not Screened	1.28 (0.87–1.89)	0.05 (0.01–0.21)	0.71 (0.39–1.29)	0.18 (0.08–0.38)	0.16 (0.05–0.50)
	Screened	2.80 (2.30–3.41)	0.19 (0.11–0.34)	1.66 (1.26–2.20)	0.33 (0.17–0.64)	0.21 (0.09–0.51)
	Difference	1.52 (0.65–2.20)	0.14 (0.001–0.22)	0.95 (0.29–1.38)	0.15 (−0.04–0.30)	0.05 (−0.12–0.19)
	% Excess	54	74	57	45	24
	Hazard Ratio	2.2 (1.43–3.4)	3.95 (0.98–15.97)	2.35 (1.32–4.19)	1.87 (0.69–5.12)	1.34 (0.4–4.49)

Notes: Tables present the adjusted cumulative incidence (breast cancer cases or breast cancer deaths per 100 individuals) through the end of follow up which occurred at breast cancer diagnosis, death, or the end of 2017. All values in parenthesis represent 95% confidence intervals. Hazard ratios compare risk of breast cancer diagnosis or death in screened and unscreened groups. In situ, localized and regional-distance incidence were derived from the SEER summary stage variable. Breast cancer death was identified from cause of death reported on death certificates. For the in-situ outcome, life expectancy was recoded into 6 month increments due to small cell sizes. For the breast cancer death outcome, life expectancy was recoded into 6-month increments and race was analyzed as non-Hispanic Black compared to all others due to small cell sizes.