



HHS Public Access

Author manuscript

Ann Intern Med. Author manuscript; available in PMC 2022 October 01.

Published in final edited form as:

Ann Intern Med. 2022 April ; 175(4): 471–478. doi:10.7326/M21-3577.

Estimation of Breast Cancer Overdiagnosis in a U.S. Breast Screening Cohort

Marc D. Ryser, PhD,

Department of Population Health Sciences, Duke University Medical Center, and Department of Mathematics, Duke University, Durham, North Carolina

Jane Lange, PhD,

Center for Early Detection Advanced Research, Knight Cancer Institute, Oregon Health Sciences University, Portland, Oregon

Lurdes Y.T. Inoue, PhD,

Department of Biostatistics, University of Washington, Seattle, Washington

Ellen S. O'Meara, PhD,

Kaiser Permanente Washington Health Research Institute, Seattle, Washington

Charlotte Gard, PhD,

Department of Economics, Applied Statistics, and International Business, New Mexico State University, Las Cruces, New Mexico

Diana L. Miglioretti, PhD,

Kaiser Permanente Washington Health Research Institute, Seattle, Washington, and Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, Davis, California

Jean-Luc Bulliard, PhD,

Centre for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland

Andrew F. Brouwer, PhD,

Corresponding Author: Marc D. Ryser, PhD, Duke University, 215 Morris Street, Durham, NC 27701; marc.ryser@duke.edu.

Author Contributions: Conception and design: R.B. Etzioni, J. Lange, M.D. Ryser. Analysis and interpretation of the data: J.

Bulliard, R.B. Etzioni, J. Lange, D.L. Miglioretti, M.D. Ryser,

Drafting of the article: R.B. Etzioni, E.S. Hwang, M.D. Ryser.

Critical revision of the article for important intellectual content: A.F. Brouwer, J. Bulliard,

R.B. Etzioni, E.S. Hwang, J. Lange, D.L. Miglioretti, M.D. Ryser.

Final approval of the article: A.F. Brouwer, J. Bulliard, R.B. Etzioni, C. Gard, E.S. Hwang, L. Inoue, J. Lange, D.L. Miglioretti, E.S. O'Meara, M.D. Ryser.

Statistical expertise: R.B. Etzioni, L. Inoue, J. Lange, D.L. Miglioretti, M.D. Ryser.

Obtaining of funding: R.B. Etzioni, D.L. Miglioretti, M.D. Ryser.

Administrative, technical, or logistic support: E.S. Hwang.

Collection and assembly of data: R.B. Etzioni, C. Gard, J. Lange, D.L. Miglioretti, E.S. O'Meara, M.D. Ryser.

Disclaimer: The statements in this report are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute and its Board of Governors or Methodology Committee, National Cancer Institute, or National Institutes of Health.

Disclosures: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M21-3577.

Reproducible Research Statement: *Study protocol and data set:* Not available. *Statistical code:* Available on https://GitHub.com/mdryser/BCSC_Overdiagnosis.

Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan

E. Shelley Hwang, MD, MPH,

Department of Surgery, Duke University Medical Center, Durham, North Carolina

Ruth B. Etzioni, PhD

Program in Biostatistics, Fred Hutchinson Cancer Research Center, Seattle, Washington

Abstract

Background: Mammography screening can lead to overdiagnosis—that is, screen-detected breast cancer that would not have caused symptoms or signs in the remaining lifetime. There is no consensus about the frequency of breast cancer overdiagnosis.

Objective: To estimate the rate of breast cancer overdiagnosis in contemporary mammography practice accounting for the detection of nonprogressive cancer.

Design: Bayesian inference of the natural history of breast cancer using individual screening and diagnosis records, allowing for nonprogressive preclinical cancer. Combination of fitted natural history model with life-table data to predict the rate of overdiagnosis among screen-detected cancer under biennial screening.

Setting: Breast Cancer Surveillance Consortium (BCSC) facilities.

Participants: Women aged 50 to 74 years at first mammography screen between 2000 and 2018.

Measurements: Screening mammograms and screen-detected or interval breast cancer.

Results: The cohort included 35 986 women, 82 677 mammograms, and 718 breast cancer diagnoses. Among all preclinical cancer cases, 4.5% (95% uncertainty interval [UI], 0.1% to 14.8%) were estimated to be nonprogressive. In a program of biennial screening from age 50 to 74 years, 15.4% (UI, 9.4% to 26.5%) of screen-detected cancer cases were estimated to be overdiagnosed, with 6.1% (UI, 0.2% to 20.1%) due to detecting indolent preclinical cancer and 9.3% (UI, 5.5% to 13.5%) due to detecting progressive preclinical cancer in women who would have died of an unrelated cause before clinical diagnosis.

Limitations: Exclusion of women with first mammography screen outside BCSC.

Conclusion: Based on an authoritative U.S. population data set, the analysis projected that among biennially screened women aged 50 to 74 years, about 1 in 7 cases of screen-detected cancer is overdiagnosed. This information clarifies the risk for breast cancer overdiagnosis in contemporary screening practice and should facilitate shared and informed decision making about mammography screening.

Primary Funding Source: National Cancer Institute.

TOC blurb

Overdiagnosis from screening can result from the detection of indolent preclinical cancer or progressive preclinical cancer where the person would have died of an unrelated cause before clinical diagnosis. This article uses statistical modeling to account for both types of overdiagnosis in estimating the rate of screen-detected breast cancer that is overdiagnosed.

Keywords

Breast cancer; Breast cancer screening; Cancer epidemiology; Cancer screening; Mammography; Natural history of disease; Overdiagnosis; Prevention, policy, and public health

It is now commonly understood that breast cancer screening is subject to overdiagnosis—that is, the mammographic detection of cancer that would not become clinically evident in the woman’s remaining lifetime. The U.S. Preventive Services Task Force cites overdiagnosis as one of the chief potential harms associated with mammography screening because of the burden and adverse consequences of unnecessary treatments (1). Therefore, knowledge about overdiagnosis is critical for supporting shared decision making, as recommended by the U.S. Preventive Services Task Force and the American Cancer Society (2, 3). However, the risk for breast cancer overdiagnosis in contemporary screening programs remains uncertain, with estimates varying from 0% to 54% (1).

This variation is attributable to differing overdiagnosis definitions, study settings, and estimation methods (4, 5). Regarding definition, some studies measure the frequency of overdiagnosis among all cases of diagnosed cancer, whereas others measure overdiagnosis only among cases of screen-detected cancer. Regarding study setting, many studies examine overdiagnosis after the introduction of population screening, which typically does not offer a counterfactual population control; others focus on screening trials, which include a control group but only provide the potential for unbiased estimation under certain trial designs (6). Finally, the analytic method used strongly influences variations in the overdiagnosis estimates, yet it is possibly the least well understood.

There are 2 main methods for estimating the frequency of overdiagnosis (5). The excess-incidence method directly acknowledges the notion that an overdiagnosed cancer is an excess cancer detected by screening that would not have been diagnosed without screening. It is an empirical approach that directly compares disease incidence among screened and unscreened populations. The lead-time method applies a mechanistic understanding of how overdiagnosis happens as a consequence of the early detection process. It uses statistical modeling to estimate the underlying disease latency and life tables to calculate the risk for death from other causes between screen detection and the end of the latent preclinical period (before symptoms or signs).

Both methods are subject to limitations. Excess incidence requires a reliable observation of disease incidence in the absence of screening; this is not always possible once screening has been broadly adopted. The validity of excess incidence as a proxy for overdiagnosis has been questioned, and the approach has been shown to generate inflated estimates except under specific conditions that are rarely met in practice (4–8). Indeed, results from excess incidence studies generally far exceed those from lead-time studies (5). The lead-time approach is more versatile but requires specifying an appropriate model of underlying disease dynamics and sourcing a sufficiently rich data set to permit estimation of model parameters (9). Historically, models of the natural history of breast cancer have posited a simple progression from healthy to preclinical and clinical disease states (10, 11). Such models have been criticized for not accommodating heterogeneity of breast cancer natural

history (12, 13). This includes the potential for breast tumors to remain indolent, particularly among preinvasive ductal carcinoma in situ (DCIS) lesions (14, 15). Accounting for a mixture of progressive and nonprogressive cancer complicates modeling and may lead to nonidentifiability, which occurs when multiple model solutions explain the observed data equally well (16, 17). Recent modeling studies that include a mixture of indolent and progressive cancer generally do not confirm that their parameters are identifiable, leading to uncertainty about the reliability of their estimates (18–20).

Here we estimate the frequency of breast cancer overdiagnosis using the lead-time method, accounting for a mixture of progressive and nonprogressive cancer and explicitly assessing model identifiability on the basis of our previous methodological work (9). To produce results that apply to U.S. population screening programs, we source screening and cancer diagnosis histories from women in the Breast Cancer Surveillance Consortium (BCSC). The BCSC is the most authoritative U.S. data source on breast cancer screening practices and outcomes (21) and the most relevant to quantify the burden of overdiagnosis in contemporary U.S. breast screening practice.

Methods

Overview

Our overdiagnosis estimates were based on individual screening and diagnosis histories from a cohort of women who received 1 or more screening mammograms at a BCSC facility between 2000 and 2018. Because overdiagnosis is not directly observable (Figure 1, *A–C*), we estimated it indirectly through Bayesian inference. We first used data on screen and interval cancer incidence (Figure 1, *D*) to estimate the underlying latency and fraction of indolent cancer. We then used these estimates together with life tables on the risk for death from causes other than breast cancer to predict the extent of overdiagnosis in a cohort of women undergoing regular mammography screening. A technical appendix complementing this section is found in Supplement 1 (available at [Annals.org](https://annals.org)).

Study Cohort

The BCSC is a racially, ethnically, and geographically diverse cohort that is representative of the mammography screening population in the United States (21, 22). The BCSC registries collect clinical and self-reported data from participating imaging facilities, including screening mammograms; diagnostic imaging, including additional imaging for work-up of a positive screen result; and benign and malignant pathology reports from breast biopsies. At each BCSC registry, breast cancer outcomes are also obtained through linkage with regional population-based SEER (Surveillance, Epidemiology, and End Results) programs and state tumor registries. Vital status is obtained through linkage to cancer registries and state death records. Each BCSC registry and the Statistical Coordinating Center have received institutional review board approval for either active or passive consenting processes or a waiver of consent and a Federal Certificate of Confidentiality and other protections for the identities of women, physicians, and facilities who are subjects of this research. All procedures were Health Insurance Portability and Accountability Act compliant.

We included women aged 50 to 74 years whose first screen occurred at a BCSC facility between 2000 and 2018 from 4 BCSC registries that are permitted to share cancer data with external, non-BCSC researchers. An additional 2 BCSC registries provided data from 2000 to 2009 before laws prohibited cancer registry data sharing. Women were excluded from the analysis if they reported having received a mammogram or a breast cancer diagnosis before their first mammogram in the BCSC registry. Mammography follow-up was censored by the diagnosis of the first breast cancer, death, 2018 (2009 for 2 registries), or a woman no longer attending a facility within the registry. Cancer follow-up was censored at death, age 89 years, 18 months past the last screening mammogram, or 2019 (2010 for 2 sites).

Screening Tests and Cancer Diagnoses

Following BCSC definitions (23), screening mammograms were defined as bilateral mammograms with an indication of “screening,” done at least 9 months after the preceding mammogram. Both invasive breast carcinoma and DCIS were counted as cancer diagnoses. A cancer was coded as screen detected if it was diagnosed within 12 months of a screening mammogram in which the final assessment (including any diagnostic work-up) resulted in a Breast Imaging-Reporting and Data System (BI-RADS) score of 3, 4, or 5, or if the final assessment was a BI-RADS score of 0 or missing and the cancer was diagnosed within 3 months (23); all other cases of cancer were coded as interval cancer. We used a 12-month interval after a BI-RADS score of 3, 4, or 5 to define a screen-detected cancer because this is consistent with the BCSC’s method of detection variable (24); more than 98% of such cases of cancer were detected within 9 months of their examination.

Estimating the Natural History of Breast Cancer

Although the exact time of onset with preclinical cancer and the period of tumor latency are not directly observable in real life (Figure 1, *A*), we can use a mathematical model together with the observed patterns of screen and interval cancer incidence (Figure 1, *D*) to estimate the underlying breast cancer dynamics (Figure 1, *E*). We constructed a model with the following components: age-dependent incidence of preclinical disease; sensitivity of screening examinations to detect a preclinical cancer; chance of a preclinical cancer to be nonprogressive; and time from onset of progressive preclinical disease to detection with clinical symptoms or signs, also known as sojourn time (Figure 1, *E*). Our estimation procedure aimed to find the most plausible values for these components, also called parameters, given the observed screening and diagnosis patterns. Because we censored deaths from causes other than breast cancer when fitting the model to the data, there was no need to explicitly model the competing risk mortality.

We used a Bayesian learning algorithm to identify the parameter values that best matched the observed data patterns. This permitted incorporating existing knowledge about model components into a set of prior beliefs, or prior distributions, on the model parameters. The prior beliefs were iteratively updated—that is, learning from the data—to provide estimates of the model parameters tailored to the data yielding posterior distributions of the model components (revised estimates of parameter values informed by observed data).

The Bayesian learning algorithm updates the model parameters using the likelihood, a function that describes how the observed data on screen and interval incidence relates to the model components. To understand how the likelihood is calculated, consider the concrete example of a woman who received 3 negative mammogram results before being diagnosed with a screen-detected cancer on the fourth screen (Figure 1, *D*: screen-detected cancer). Although the cancer certainly began before the fourth screen, it may have occurred between the third and fourth screen, or it may have started before the third screen and been undetected on 1 or more false-negative screens. Adding up these possibilities across all women allows us to “learn” the parameter combinations that best fit the data. To ensure that parameter selection was primarily guided by the information contained in the data, we used noninformative priors for all parameters except for the screening test sensitivity, which has extensive existing published studies to estimate an informative prior probability distribution that helps to identify a unique best-fitting combination of model parameters.

Uncertainty about the model parameters was captured by 95% credible intervals (CrIs) around the posterior mean, analogous to classical CIs around the point estimate. A benefit of the Bayesian approach is that it naturally propagates uncertainty around the parameter estimates to uncertainty around the predicted rate of overdiagnosis.

Overdiagnosis Prediction

We defined the rate of breast cancer overdiagnosis as the proportion of screen-detected cancer that was either nonprogressive or progressive but would not have progressed to clinical disease before the woman died of causes unrelated to breast cancer (Figure 1, *B*). To estimate this proportion, we considered a cohort of women whose parameters for disease natural history were given by the best-fitting parameter combinations and who had regular (annual or biennial) screening, starting at age 50 years and until age 74 years or death from a cause unrelated to breast cancer, whichever occurred first. We modeled the competing mortality risk on the basis of a published age-cohort model (25) for a 1971 birth cohort. Residual uncertainty around the predicted mean rate of overdiagnosis was captured by 95% prediction intervals (PIs).

Sensitivity Analyses

We examined the overdiagnosis predictions’ sensitivity to alternative prior distributions for the sojourn time and the screening test sensitivity. We also assessed the overdiagnosis rate under a purely progressive disease model that did not allow for nonprogressive cancer.

Role of the Funding Source

This study was funded by the National Cancer Institute, which had no role in design, analysis, manuscript preparation, or the decision to submit the manuscript for publication.

Results

Screening Cohort Summary

Our analysis included 35 986 women who had mammographic screening (Table 1). Among women of known race, 64.4% were White, 19.0% Asian, and 12.1% Black. Among women

of known ethnicity, 11.0% identified as Hispanic. Median age at first screen was 56 years, with a local peak at 65 years, corresponding to the starting age of Medicare eligibility (Figure 2, *left*). On average, each woman received 2.3 screening tests (range, 1 to 17), for a total of 82 677 screening tests across the cohort. Most (51.3%) had a single test, and 92.1% received 5 or fewer tests. The time between consecutive screens indicated primarily annual and biennial screening patterns (Figure 2, *middle*). A total of 718 cases of breast cancers were diagnosed; among the 671 cases of known type, 79.0% were invasive cancers and 21.0% were DCIS. Most cases of cancer (90.0%) were screen detected, and the highest number of cancer cases were diagnosed during the first screening round (Figure 2, *right*).

Natural History of Breast Cancer

We estimated 5 model parameters (Figure 1, *E*; Table 2). The incidence of preclinical cancer increased with age, and 4.5% (95% CrI, 0.1% to 14.8%) of all cases of preclinical cancer were nonprogressive. The mean tumor latency among cases of progressive cancer was estimated to be 6.6 years (CrI, 4.9 to 8.6), and the screening test sensitivity 81.4% (CrI, 73.1% to 88.7%).

Overdiagnosis Between Ages 50 and 74 Years

For a biennial screening program for persons between the ages of 50 and 74 years, the overall predicted overdiagnosis rate among screen-detected cancer cases was 15.4% (95% PI, 9.4% to 26.5%). The overdiagnosis rate increased from 11.5% (PI, 3.8% to 28.3%) at the first screen at age 50 years to 23.6% (PI, 17.7% to 31.9%) at the last screen at age 74 years (Figure 3 and Table 3). The wide PIs indicate uncertainty around the estimates but also identify the most likely ranges. For instance, the posterior probability of an overall overdiagnosis rate between 5% and 20% was 82.6%, whereas lower (<10%) and higher (>20%) rates were less likely, with posterior probabilities of 4.8% and 12.6%, respectively. The relative contributions to overdiagnosis from the detection of nonprogressive and progressive types of preclinical cancer differed by screening round (Figure 3). Overall, the rate of overdiagnosis among screen-detected cancer cases that was due to nonprogressive cancer decreased from 8.4% (PI, 0.3% to 26.4%) at the first screen to 5.5% (PI, 0.2% to 17.0%) at the last screen (Table 3). Conversely, the rate of overdiagnosis among screen-detected cancer cases that was due to progressive disease that would not have become clinically evident before death from other causes increased from 3.1% (PI, 1.6% to 5.1%) at the first screen to 18.1% (PI, 11.9% to 24.5%) at the last screen.

Doubling the screening frequency from biennial to annual yielded an overall predicted overdiagnosis rate of 14.6% (PI, 9.4% to 23.9%), with patterns similar to those under biennial screening (Figure 4 of Supplement 2, available at [Annals.org](https://www.annals.org)).

Sensitivity Analyses

The overdiagnosis rate was insensitive to alternative prior distributions for the screening test sensitivity (Figure 5 of Supplement 2, available at [Annals.org](https://www.annals.org)). When using informative priors for the sojourn time based on published consensus estimates, changes in the predicted overdiagnosis rate were negligible for all but the most restrictive prior distributions (Figure 6 of Supplement 2, available at [Annals.org](https://www.annals.org)). Replacing the mixture model with a purely

progressive disease model led to a lower predicted overdiagnosis rate under biennial screening of 10.9% (PI, 8.0% to 15.0%) (Table 2 of Supplement 2, available at [Annals.org](#)).

Discussion

In this study, we sought to produce a reliable estimate of overdiagnosis that is relevant to contemporary U.S. practice by using an authoritative data source and a modeling approach that overcome key limitations of many previously published studies. For a biennial screening program for persons between the ages of 50 and 74 years, which corresponds to the U.S. Preventive Services Task Force recommendation for average risk women (2), we estimated that approximately 1 in 7 screen-detected cases would be overdiagnosed. Because the estimated tumor latency was considerably greater than 2 years, more frequent screening had a marginal effect on the number of screen-detected cancers overall.

Among all overdiagnosed cancers in the biennial screening program for persons between the ages of 50 and 74 years, one third was due to the detection of nonprogressive cancer and two thirds were due to the detection of progressive cancer that would not have progressed to clinical disease before the woman died of other causes. Such a decomposition of overdiagnosis into its principal sources is rarely reported, and it allows for a more informative characterization of age-dependent patterns of overdiagnosis.

A comparison of our estimates against those from other studies is not straightforward because of differences in overdiagnosis definitions and screening practices. Our overdiagnosis estimate is generally lower than those from similar excess-incidence studies, which are prone to overestimation (6). For instance, the Canadian screening trial estimated overdiagnosis rates of 30% (26) among screen-detected cancers, as compared with our estimate of 15%. In a population study, Bleyer and Welch (27) estimated that 31% of all diagnosed breast cancer cases were overdiagnosed, which corresponds to an even higher overdiagnosis rate among cases of screen-detected cancer (28). For context, our model deems overdiagnosis rates of this magnitude unlikely, with posterior probabilities of 3% and 1% for overdiagnosis rates exceeding 25% and 30%, respectively.

On the other hand, our estimate of overdiagnosis is higher than those of previous modeling studies, ranging from 1% to 12% (19, 29–31). For context, our model estimates overdiagnosis rates below 10% to be unlikely, with a posterior probability of 5%. The discrepancy between our findings and previous modeling studies is likely because of differences in screening and diagnostic practices in the study populations as well as in model assumptions. For instance, we found that not modeling cases of indolent cancer led to lower predicted overdiagnosis rates.

Our work adds to the collection of modeling studies that have augmented the purely progressive disease model by adding nonprogressive cases (18–20, 32–36). Fryback and colleagues (32) did not assess overdiagnosis but estimated that as many as 42% of incident tumors could be of limited malignant potential, defined as small lesions that regressed if not detected within 2 years. On the basis of French registry data, Seigneurin and colleagues (20) estimated that approximately 7% of cases of screen-detected cancer were nonprogressive,

which is similar to our estimate. Using data from the biennial Stockholm breast screening program, Wu and colleagues (19) estimated that less than 1% of screen-detected cases were overdiagnosed (19). Importantly, none of the previous modeling studies allowing for nonprogressive lesions performed identifiability analyses to ensure uniqueness of their estimates.

Because DCIS lesions have a limited propensity to progress, it is possible that they accounted for a large fraction of overdiagnosed lesions in our cohort. Indeed, with DCIS accounting for 20% of screen-detected lesions and 70% to 80% expected to be nonprogressive or very slowly growing (14, 37), this implies an overdiagnosis rate of approximately 15% among all cases of screen-detected cancer solely because of DCIS diagnoses. Because this estimate coincides with our overall overdiagnosis estimate, we hypothesize that overdiagnosis may be infrequent among cases of invasive cancer. However, a formal assessment of this hypothesis was not feasible because we did not estimate separate natural histories for DCIS and invasive cancers because of identifiability concerns.

Model-based analyses are subject to different sources of uncertainty, including model and parameter uncertainty (38). Model uncertainty is introduced when making specific choices about the mathematical structure of the natural history model and the form of the prior distributions. We did a range of sensitivity analyses that confirmed the overdiagnosis predictions were largely insensitive to variations in both model structure and prior distributions. Parameter uncertainty in turn captures residual uncertainty about the true parameter values and arises from fitting the model to real-world data of limited information content. Our Bayesian modeling approach naturally accounted for parameter uncertainty by propagating it through to the predicted frequency of overdiagnosis.

Differences between our parameter estimates and those in previously published studies can be explained by contextual factors. Our inferred screening test sensitivity of 81% characterizes the sensitivity of the full screening episode, including screening mammogram, diagnostic work-up, and biopsy referral. Although influenced by the informative prior distribution, this parameter estimate was updated by the information contained in the data. Our estimate of 6.6 years for the mean sojourn time of progressive cancers is higher than a previous consensus estimate of 2 to 4 years based on data from older screening trials (7), yet remains within the range of previous modeling studies in which estimates were as high as 7.5 years (39). The discrepancy in detectable latency estimates compared with older studies is likely due to advances in screening technologies and our inclusion of DCIS tumors. Advances in mammography technology and supplemental screening among women with dense breasts (40) could have rendered lesions detectable at smaller sizes, which would have lengthened the sojourn time.

Naturally, our work is subject to limitations stemming from both data and method. Although the BCSC data is the largest, highest-quality U.S.-based data resource available for breast screening research, it is encounter based and therefore not able to distinguish persons who are lost to follow-up from those who have not yet come back for their next screening examination. To address this limitation, we censored cancer diagnoses 18 months after the woman's last screening examination in the data. Second, knowledge about each woman's

full screening history is needed for valid inference of natural history. Consequently, we were limited to including only those women who had their first screening examination within the BCSC. This led to event numbers that were too low to fit models accounting for patient features (such as race, ethnicity, or breast density), age-specific sojourn times, and histologic or molecular tumor subtypes. Finally, we acknowledge the potential for model misspecification but note that the model's predicted risk for developing breast cancer between the ages of 50 and 80 years (8.5%) matches that in the U.S. population (8.8%) (41).

In conclusion, we find that the rate of overdiagnosis in U.S. population-based mammography screening is unlikely to be as high as suggested by prominent excess-incidence studies. We hope that our findings will bring the field closer to a consensus estimate and facilitate decision making about mammography screening. Our estimates of the frequency and the age dependence of overdiagnosis can be provided along with information about false-positive rates to balance estimates of mammography screening benefits as part of a process of shared and informed decision making.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment:

The authors thank the participating women, mammography facilities, and radiologists for the data they have provided.

Financial Support:

By grants R00CA207872, R01CA242735, and R01CA192492 from the National Institutes of Health. Data collection was additionally supported by the BCSC (www.bcsc-research.org/) with funding from the National Cancer Institute (grants P01CA154292 and U54CA163303) and the Patient-Centered Outcomes Research Institute (grant PCS-1504-30370). Cancer and vital status data collection were supported in part by several U.S. state public health departments and cancer registries (www.bcsc-research.org/about/work-acknowledgement).

References

1. Nelson HD, Pappas M, Cantor A, et al. Harms of breast cancer screening: systematic review to update the 2009 U.S. Preventive Services Task Force recommendation. *Ann Intern Med* 2016;164:256–67. doi:10.7326/M15-0970 [PubMed: 26756737]
2. Siu AL; U.S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2016;164:279–96. doi:10.7326/M15-2886 [PubMed: 26757170]
3. Oeffinger KC, Fontham ET, 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;314:1599–614. doi:10.1001/jama.2015.12783 [PubMed: 26501536]
4. Biesheuvel C, Barratt A, Howard K, et al. Effects of study methods and biases on estimates of invasive breast cancer overdiagnosis with mammography screening: a systematic review. *Lancet Oncol* 2007;8:1129–1138. doi:10.1016/S1470-2045(07)70380-7 [PubMed: 18054882]
5. Etzioni R, Gulati R, Mallinger L, et al. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. *Ann Intern Med* 2013;158:831–8. doi:10.7326/0003-4819-158-11-201306040-00008 [PubMed: 23732716]
6. Gulati R, Feuer EJ, Etzioni R. Conditions for valid empirical estimates of cancer overdiagnosis in randomized trials and population studies. *Am J Epidemiol* 2016;184:140–7. doi:10.1093/aje/kwv342 [PubMed: 27358266]

7. Duffy SW, Parmar D. Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time. *Breast Cancer Res* 2013;15:R41. doi:10.1186/bcr3427 [PubMed: 23680223]
8. Shen Y, Zelen M. Screening sensitivity and sojourn time from breast cancer early detection clinical trials: mammograms and physical examinations. *J Clin Oncol* 2001;19:3490–9. [PubMed: 11481355]
9. Ryser MD, Gulati R, Eisenberg MC, et al. Identification of the fraction of indolent tumors and associated overdiagnosis in breast cancer screening trials. *Am J Epidemiol* 2019;188:197–205. doi:10.1093/aje/kwy214 [PubMed: 30325415]
10. Shen Y, Zelen M. Parametric estimation procedures for screening programmes: stable and nonstable disease models for multimodality case finding. *Biometrika* 1999;86:503–515. doi:10.1093/biomet/86.3.503
11. Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* 1984;40:1–14. [PubMed: 6733223]
12. Baker SG, Prorok PC, Kramer BS. Lead time and overdiagnosis [Editorial]. *J Natl Cancer Inst* 2014;106. doi:10.1093/jnci/dju346
13. Zahl PH, Jørgensen KJ, Gøtzsche PC. Lead-time models should not be used to estimate overdiagnosis in cancer screening. *J Gen Intern Med* 2014;29:1283–6. doi:10.1007/s11606-014-2812-2 [PubMed: 24590736]
14. Ryser MD, Weaver DL, Zhao F, et al. Cancer outcomes in DCIS patients without locoregional treatment. *J Natl Cancer Inst* 2019;111:952–960. doi:10.1093/jnci/djy220 [PubMed: 30759222]
15. van Seijen M, Lips EH, Thompson AM, et al. ; PRECISION team. Ductal carcinoma in situ: to treat or not to treat, that is the question. *Br J Cancer* 2019;121:285–292. doi:10.1038/s41416-019-0478-6 [PubMed: 31285590]
16. Raue A, Kreutz C, Maiwald T, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 2009;25:1923–9. doi:10.1093/bioinformatics/btp358 [PubMed: 19505944]
17. Brouwer AF, Meza R, Eisenberg MC. Parameter estimation for multistage clonal expansion models from cancer incidence data: a practical identifiability analysis. *PLoS Comput Biol* 2017;13:e1005431. doi:10.1371/journal.pcbi.1005431 [PubMed: 28288156]
18. Alagoz O, Ergun MA, Cevik M, et al. The University of Wisconsin Breast Cancer Epidemiology Simulation Model: an update. *Med Decis Making* 2018;38:99S–111S. doi:10.1177/0272989X17711927 [PubMed: 29554470]
19. Wu WY, Törnberg S, Elfström KM, et al. Overdiagnosis in the population-based organized breast cancer screening program estimated by a non-homogeneous multi-state model: a cohort study using individual data with long-term follow-up. *Breast Cancer Res* 2018;20:153. doi:10.1186/s13058-018-1082-z [PubMed: 30558679]
20. Seigneurin A, François O, Labarère J, et al. Overdiagnosis from non-progressive cancer detected by screening mammography: stochastic simulation study with calibration to population based registry data. *BMJ* 2011;343:d7017. doi:10.1136/bmj.d7017 [PubMed: 22113564]
21. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *AJR Am J Roentgenol* 1997;169:1001–8. [PubMed: 9308451]
22. Lehman CD, Arao RF, Sprague BL, et al. National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium. *Radiology* 2017;283:49–58. doi:10.1148/radiol.2016161174 [PubMed: 27918707]
23. Breast Cancer Surveillance Consortium. Standard definitions Accessed at www.bsc-research.org/data/bcsc_standard_definitions on 23 December 2021.
24. Sprague BL, Miglioretti DL, Lee CI, et al. New mammography screening performance metrics based on the entire screening episode. *Cancer* 2020;126:3289–3296. doi:10.1002/cncr.32939 [PubMed: 32374471]
25. Gangnon RE, Stout NK, Alagoz O, et al. Contribution of breast cancer to overall mortality for US women. *Med Decis Making* 2018;38:24S–31S. doi:10.1177/0272989X17717981 [PubMed: 29554467]

26. Baines CJ, To T, Miller AB. Revised estimates of overdiagnosis from the Canadian National Breast Screening Study. *Prev Med* 2016;90:66–71. doi:10.1016/j.ypmed.2016.06.033 [PubMed: 27374944]
27. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med* 2012;367:1998–2005. doi:10.1056/NEJMoa1206809 [PubMed: 23171096]
28. Etzioni R, Xia J, Hubbard R, et al. A reality check for overdiagnosis estimates associated with breast cancer screening. *J Natl Cancer Inst* 2014;106. doi:10.1093/jnci/dju315
29. de Gelder R, Fracheboud J, Heijnsdijk EA, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. *Prev Med* 2011;53:134–40. doi:10.1016/j.ypmed.2011.06.009 [PubMed: 21718717]
30. Gunsoy NB, Garcia-Closas M, Moss SM. Estimating breast cancer mortality reduction and overdiagnosis due to screening for different strategies in the United Kingdom. *Br J Cancer* 2014;110:2412–9. doi:10.1038/bjc.2014.206 [PubMed: 24762956]
31. Bulliard JL, Beau AB, Njor S, et al. Breast cancer screening and overdiagnosis. *Int J Cancer* 2021. doi:10.1002/ijc.33602
32. Fryback DG, Stout NK, Rosenberg MA, et al. The Wisconsin Breast Cancer Epidemiology Simulation Model. *J Natl Cancer Inst Monogr* 2006;37–47. [PubMed: 17032893]
33. Schechter CB, Near AM, Jayasekera J, et al. Structure, function, and applications of the Georgetown–Einstein (GE) Breast Cancer Simulation Model. *Med Decis Making* 2018;38:66S–77S. doi:10.1177/0272989X17698685 [PubMed: 29554462]
34. van den Broek JJ, van Ravesteyn NT, Heijnsdijk EA, et al. Simulating the impact of risk-based screening and treatment on breast cancer outcomes with MISCAN-Fadia. *Med Decis Making* 2018;38:54S–65S. doi:10.1177/0272989X17711928 [PubMed: 29554469]
35. Lee SJ, Li X, Huang H, et al. The Dana-Farber CISNET model for breast cancer screening strategies: an update. *Med Decis Making* 2018;38:44S–53S. doi:10.1177/0272989X17741634 [PubMed: 29554465]
36. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;380:1778–86. doi:10.1016/S0140-6736(12)61611-0 [PubMed: 23117178]
37. Erbas B, Provenzano E, Armes J, et al. The natural history of ductal carcinoma in situ of the breast: a review. *Breast Cancer Res Treat* 2006;97:135–44. [PubMed: 16319971]
38. Spiegelhalter DJ, Riesch H. Don't know, can't know: embracing deeper uncertainties when analysing risks. *Philos Trans A Math Phys Eng Sci* 2011;369:4730–50. doi:10.1098/rsta.2011.0163 [PubMed: 22042895]
39. Weedon-Fekjaer H, Vatten LJ, Aalen OO, et al. Estimating mean sojourn time and screening test sensitivity in breast cancer mammography screening: new results. *J Med Screen* 2005;12:172–8. [PubMed: 16417693]
40. Huang S, Houssami N, Brennan M, et al. The impact of mandatory mammographic breast density notification on supplemental screening practice in the United States: a systematic review. *Breast Cancer Res Treat* 2021;187:11–30. doi:10.1007/s10549-021-06203-w [PubMed: 33774734]
41. National Cancer Institute. SEER*Explorer Accessed at <https://seer.cancer.gov/explorer/> on 23 December 2021.

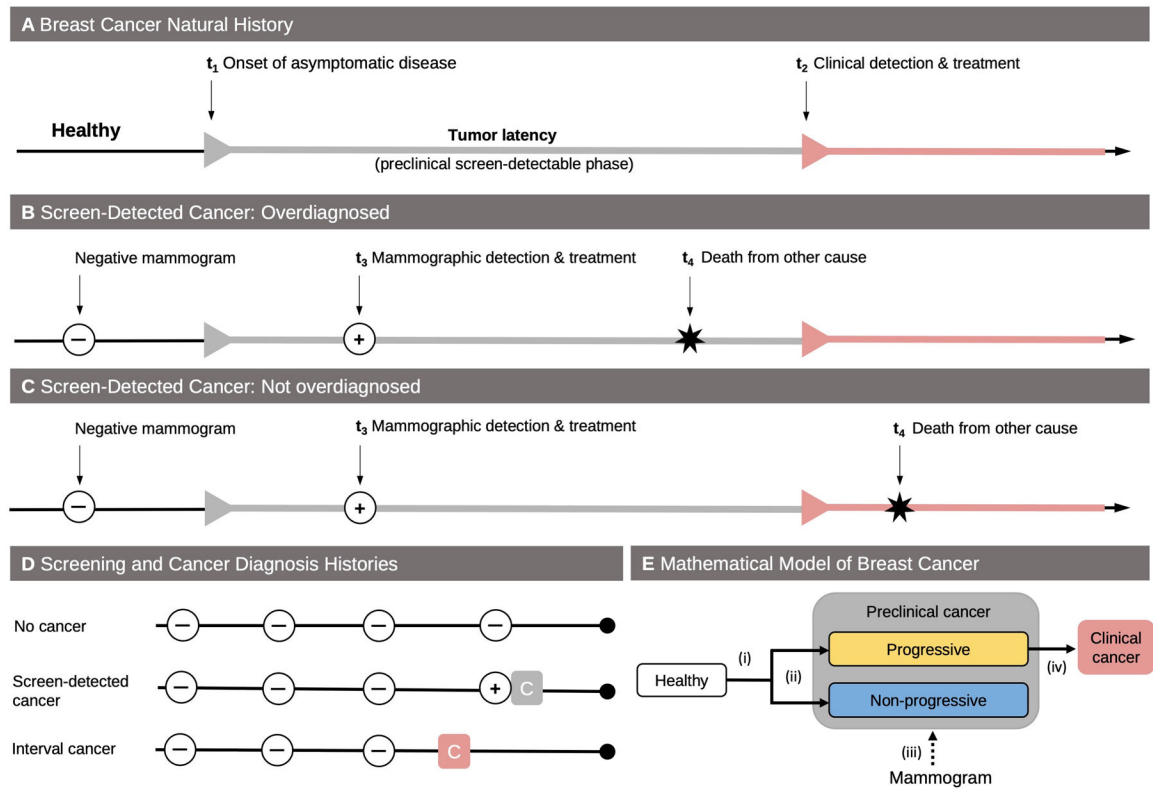


Figure 1. Overdiagnosis definition, data structure, and mathematical model. **A.** In the absence of screening, healthy women can develop preclinical cancer (t_1). After a period of asymptomatic but screen-detectable tumor latency (sojourn time), women with preclinical cancer develop clinical cancer (t_2). **B.** A screen-detected cancer at age t_3 is overdiagnosed if, in the absence of screening, death from a breast cancer unrelated cause (age t_4) would have occurred before onset with symptoms or signs ($t_4 < t_3$). **C.** On the other hand, the detected cancer is not overdiagnosed if the time of death would have occurred after onset with symptoms or signs ($t_4 > t_3$). **D.** The screening and cancer diagnosis histories of women in the Breast Cancer Surveillance Consortium cohort can be assigned to 1 of 3 groups: no cancer diagnosis, diagnosis with a screen-detected cancer (gray square), and diagnosis with an interval cancer (red square). **E.** The mathematical model used for natural history estimation and overdiagnosis prediction has several components: (i) age-dependent incidence of preclinical cancer, (ii) fraction of preclinical cancer that is nonprogressive, (iii) sensitivity of screening tests, and (iv) tumor latency among progressive cancer.

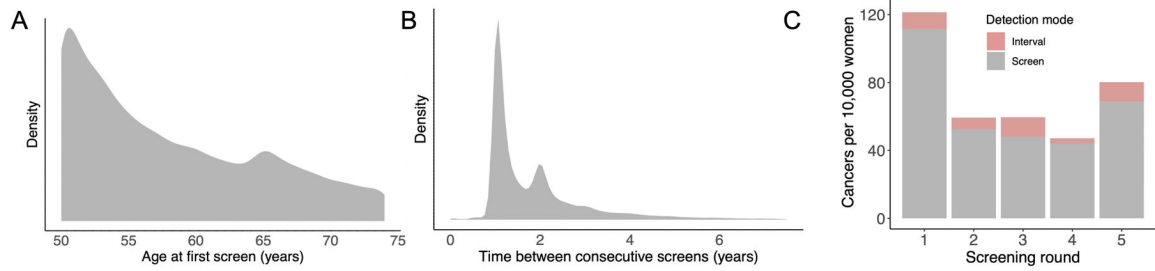


Figure 2.

Data summary. In the middle and right panels, Gaussian kernel density estimates were used to smooth the data (see Supplement 1). **Left.** Participant age at first screen in the analyzed cohort. **Middle.** Time between consecutive screens, across all participants. **Right.** Cases of cancers (per 10 000 women) diagnosed during the first 5 screening rounds, across all participants and by method of detection.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

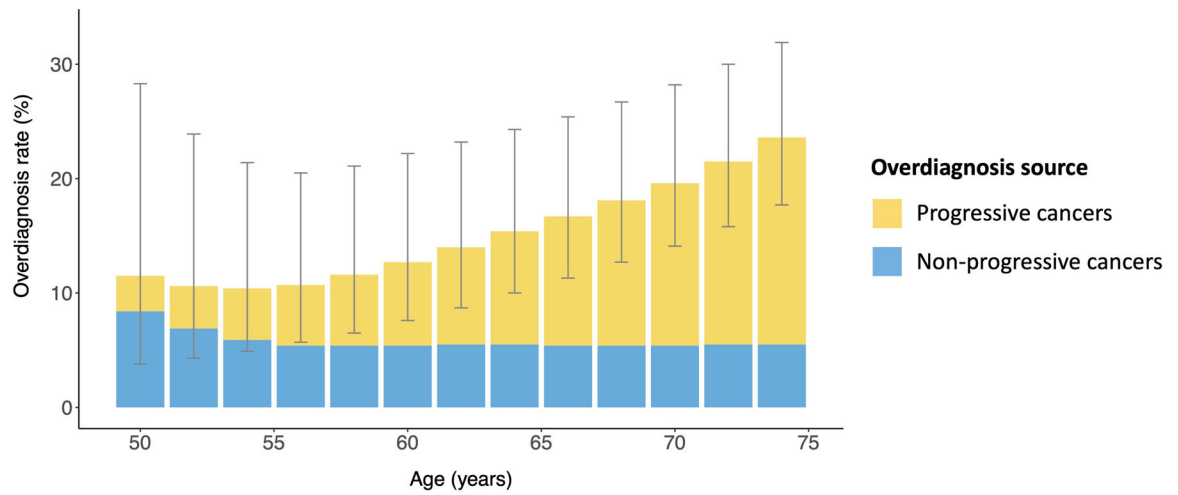


Figure 3. Overdiagnosis in women undergoing biennial screening, ages 50 to 74 years. For women undergoing biennial screening starting at age 50 years, the mean predicted overdiagnosis rate is shown by the height of each bar for each screen until age 74 years (gray lines represent the 95% prediction intervals). There are 2 sources of overdiagnosis: the detection of progressive preclinical cancer that would not have progressed to clinical cancer before death from a breast cancer unrelated cause (white) and the detection of nonprogressive preclinical cancer (green).

Table 1.

Screening Cohort Summary

Characteristics	Value
Screening participants (<i>n</i> = 35 986)	
Median age at first screen (IQR) [range], <i>y</i>	56 (52–64) [50–74]
Race, <i>n</i> (%)	
White	20 001 (55.6)
Asian	5897 (16.4)
Black	3754 (10.4)
Other	1417 (3.9)
Unknown	4917 (13.7)
Ethnicity, <i>n</i> (%)	
Hispanic	3704 (10.3)
Non-Hispanic	29 964 (83.3)
Unknown	2318 (6.4)
Screens (<i>n</i> = 82 677)	
Screens per participant, <i>n</i>	
Mean (range)	2.3 (1–17)
Median (IQR)	1 (1–3)
Number of screens, <i>n</i> (%) of participants	
1	18 451 (51.3)
2	7114 (19.8)
3	3848 (10.7)
4	2211 (6.1)
5	1516 (4.2)
6	2846 (7.9)
Cancer (<i>n</i> = 718)	
Mode of detection, <i>n</i> (%)	
Screen detected	645 (90.0)
Interval	73 (10.0)
Cancer type, <i>n</i> (%)	
Invasive	530 (73.8)
In situ	141 (19.6)
Unknown	47 (6.5)

IQR = interquartile range.

Table 2.

Parameter Estimates

Model Parameter	Posterior Mean (95% Credible Interval)
Preclinical onset rate for ages [40, 55), year ⁻¹	0.0017 (0.0015–0.0021)
Preclinical onset rate for ages [55, 65), year ⁻¹	0.0029 (0.0024–0.0033)
Preclinical onset rate for ages [65, ∞), year ⁻¹	0.0035 (0.0029–0.0042)
Mean sojourn time, y	6.6 (4.9–8.6)
Fraction indolent cancers, %	4.5 (0.1–14.8)
Screening episode sensitivity, %	81.4 (73.1–88.7)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Predicted Overdiagnosis Rates for a Biennial Screening Program, Ages 50 to 74 Years

Overdiagnosis Measure	Mean Predicted Overdiagnosis Rate (95% Prediction Interval), %
Overall	
Total	15.4 (9.4–26.5)
Indolent cancer contribution	6.1 (0.2–20.1)
Progressive cancer contribution	9.3 (5.5–13.5)
At first screen (age 50 y)	
Total	11.5 (3.8–28.3)
Indolent cancer contribution	8.4 (0.3–26.4)
Progressive cancer contribution	3.1 (1.6–5.1)
At fifth screen (age 58 y)	
Total	11.6 (6.5–21.1)
Indolent cancer contribution	5.4 (0.2–16.8)
Progressive cancer contribution	6.2 (3.7–9.5)
At ninth screen (age 66 y)	
Total	16.7 (11.3–25.4)
Indolent cancer contribution	5.4 (0.2–16.8)
Progressive cancer contribution	11.3 (7.1–16.0)
At 13th screen (age 74 y)	
Total	23.6 (17.7–31.9)
Indolent cancer contribution	5.5 (0.2–17.0)
Progressive cancer contribution	18.1 (11.9–24.5)