# Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies.

Mandelblatt JS, Stout NK, Schechter CB, van den Broek JJ, Miglioretti DL, Krapcho M, Trentham-Dietz A, Munoz D, Lee SJ, Berry DA, van Ravesteyn NT, Alagoz O, Kerlikowske K, Tosteson AN, Near AM, Hoeffken A, Chang Y, Heijnsdijk EA, Chisholm G, Huang X, Huang H, Ergun MA, Gangnon R, Sprague BL, Plevritis S, Feuer E, de Koning HJ, Cronin KA.

*Annals of Internal Medicine. 2016 Feb 16;164(4):215-25.*

## ABSTRACT

### Background

Controversy persists about optimal mammography screening strategies.

### Objective

To evaluate screening outcomes, taking into account advances in mammography and treatment of breast cancer.

### Design

Collaboration of six simulation models using national data on incidence, digital mammography performance, treatment effects, and other-cause mortality.

### Setting and Patients

The average-risk US female population and sub-groups with varying risk, breast density, or comorbidity.

### Setting

Unites States

### Patients

Average-risk U.S. female population and subgroups with varying risk, breast density, or comorbidity

### Interventions

Eight strategies differing by age at which screening starts (40, 45, 50 years) and screening interval (annual, biennial, and hybrid [annual in the 40s and biennial thereafter]); all strategies assumed 100% adherence and stopped at age 74.

### Measurements

Benefits (breast cancer-specific mortality reduction, breast cancer deaths averted, life-years and quality-adjusted life years); number of mammograms used; harms (false-positive results, benign biopsies, and overdiagnosis); and ratios of harms (or use) and benefits (efficiency) per 1000 screens.

### Results

Biennial strategies were consistently the most efficient for average-risk women. Biennial screening from ages 50-74 avoided a median of 7 breast cancer deaths vs. no screening; annual screening from ages 40-74 years avoided an additional 3 deaths, but yielded

1988 more false-positives and 7 more overdiagnoses per 1,000 women screened. Annual screening from ages 50-74 was inefficient (similar benefits but more harms than other strategies). For groups with a 2- to 4-fold increased risk, annual screening from age 40 had similar harms and benefits as screening average-risk women biennially from 50-74. For groups with moderate or severe comorbidity, screening could stop at age 66 to 68 years.

### Limitations

Other imaging technologies, polygenic risk, and nonadherence were not considered.

### Conclusion

Biennial screening for breast cancer is efficient for average-risk populations. Decisions regarding starting ages and intervals will ultimately depend on population characteristics and the decision-makers' weight given to the harms and benefits of screening.

### Primary Funding Source

National Institutes of Health

## INTRODUCTION

Despite decades of mammography screening for early breast cancer detection, there is no consensus on optimal strategies, target populations, or the magnitude of harms and benefits.(1-11) The 2009 US Preventive Services Task Force recommended biennial film mammography from ages 50-74, and suggested shared decision-making about screening in the 40's.(12) Since that recommendation was formulated, there have been some new data regarding screening benefits,(2,6,8,9,11,13,14) digital mammography has essentially replaced plain film,(15) and increasingly effective breast cancer systemic treatment regimens have become standard.(16) There has also been growing interest in consumer preferences and personalized screening approaches.(17-20). These factors could each affect the outcomes of breast cancer screening programs and/or alter policy decisions about population screening strategies.(17)

Modeling can inform screening policy decisions since it uses the best available evidence to evaluate a wide range of strategies, while holding selected conditions (e.g., treatment effects) constant, facilitating strategy comparisons.(21,22) Modeling also provides a quantitative summary of outcomes in different groups and assesses how preferences affect results. Collaboration of several models provides a range of plausible effects and illustrates the impact of differences in model assumptions on results.(1,7,23)

We used six well-established simulation models to synthesize current data to examine the outcomes of digital mammography screening at various starting ages and intervals among average-risk women. We also examined how breast density, risk, or comorbidity levels affect results, and whether preferences for health states related to screening and its downstream consequences affected conclusions.

## METHODS

### Strategies

We evaluated eight strategies that varied by starting age (40, 45, 50) and interval (annual, biennial, and hybrid [annual in the 40's and biennial thereafter]); all strategies stop screening at age 74. We included "no screening" as a baseline.

### Model Descriptions

The models used to evaluate the screening strategies were developed within the Cancer Intervention and Surveillance Modeling Network (CISNET) (24-30) and the research was institutional review board approved. The models included model D (Dana-Farber Cancer Institute, Boston, Massachusetts), model E (Erasmus Medical Center, Rotterdam, the Netherlands), model GE (Georgetown University Medical Center, Washington, DC and

Albert Einstein College of Medicine, Bronx, New York), model M (MD Anderson Cancer Center, Houston, Texas), model S (Stanford University, Stanford, California), and model W (University of Wisconsin, Madison, Wisconsin and Harvard Medical School, Boston, Massachusetts).

Since earlier analyses,(1) the models have undergone substantial revision to reflect advances in breast cancer control, including portrayal of molecular subtypes based on estrogen receptor (ER) and human epidermal growth factor-2 receptor (HER2) status;(23) current population incidence (31) and competing non-breast cancer mortality; digital screening; and the most current therapies.(32) All models except model S include ductal carcinoma in-situ (DCIS).

The general modeling approach is summarized below; full details including approach, construction, data sources, assumptions, and implementation are available at: https://resources.cisnet.cancer.gov/registry and at (33). Additional information is available on request and the models are available for use via collaboration.

The models begin with estimates of breast cancer incidence (31) and ER/HER2-specific survival trends *without* screening or adjuvant treatment and then overlay data on screening and molecular subtype-specific adjuvant treatment to generate observed US population incidence and breast cancer-specific mortality trends.(1,7,17,23,34) Breast cancers have a distribution of preclinical screen-detectable periods (sojourn time) and clinical detection points. Digital mammography performance characteristics depend on age, first vs. subsequent screen, time since last mammogram, and breast density. ER/HER2 status is assigned at diagnosis based on stage and age. Molecular subtype- and stage-specific treatment reduces the hazard of breast cancer death (models D, GE, M, and S) or results in a cure for some cases (models E and W). Women can die of breast cancer or other causes. Screen detection of cancer during the preclinical screen-detectable period can result in the identification (and treatment) of earlier-stage or smaller tumors than might occur via clinical detection, with a corresponding reduction in breast cancer mortality.

We used a cohort of women born in 1970 with average-risk and average breast density and follow them from age 25 (since breast cancer is rare before this age [0.08% of cases]) until death or age 100.

## Model Input Parameters

The models used a common set of age-specific variables for breast cancer incidence, digital mammography performance, treatment effects, and average and comorbidity-level specific-non-breast cancer causes of death.(20,33,35) The parameter values are available at: http://www.uspreventiveservicestaskforce.org/Page/Document/modeling-report-collaborative-modeling-of-us-breast-cancer-1/breast-cancer-screening1.(33) In addition, each group included model-specific inputs (or intermediate outputs) to represent preclinical detectable times, lead-time, and age- and ER/HER2-specific stage

distribution in screen- vs. non-screen-detected women on the basis of their specific model structure.(1,7,23-30) These model-specific parameters were based on assumptions about combinations of values that reproduced US trends in incidence and breast cancer-specific mortality, including proportions of DCIS that were nonprogressive and would not be detected without screening. Models M and W also assumed some small nonprogressive invasive cancers. The models adopted an age-period-cohort modeling approach to project breast cancer incidence rates in the absence of screening;(31,36) Model M used 1975-79 SEER rates. The models assumed 100% adherence to screening and receipt of the most effective treatment to isolate the effect of varying screening strategies.

Four models used age-specific digital mammography sensitivity values observed in the Breast Cancer Surveillance Consortium (BCSC) for detection of invasive and DCIS cancers combined (model S only uses data for invasive cancers). Separate values were used for initial and subsequent mammography by screening interval using standard BCSC definitions: annual includes data from screens occurring within 9-18 months of the prior screen and biennial includes data on screens within 19-30 months.(37,38) Model D used these data as input variables (28) and models GE, S, and W used the data for calibration.(24,25,27) Models E and M fit estimates from the BCSC and other data.(26,29)

Women with ER-positive tumors received five years of hormonal therapy and an anthracycline-based regimen accompanied by a taxane. Women with ER-negative invasive tumors received anthracycline-based regimens with a taxane. Those with HER2-positive tumors also received trastuzumab. Women with ER-positive DCIS received hormonal therapy.(16) Treatment effectiveness was based on clinical trials and was modeled as a reduction in breast cancer-specific mortality risk or increase in the proportion cured compared to ER/HER2-specific survival in the absence of adjuvant treatment.(32)

### Benefits

Screening benefits (vs. no screening or incremental to other strategies) included percent breast cancer mortality reduction, breast cancer deaths averted, and life-years (LYs) and quality-adjusted life-years (QALYs) gained because of averted or delayed breast cancer death. Benefits (and harms) were accumulated from ages 40-100 years to capture the lifetime impact of screening.

We considered preferences, or utilities to account for morbidity from screening and treatment. A disutility for age- and gender-specific general population health was first applied to quality-adjust life years.(39) These were further adjusted to account for additional decrements in life years related to undergoing screening (-0.006 for one week), evaluation of a positive screen (-0.105 for five weeks), undergoing initial treatment by stage (for the first 2 years after diagnosis), and experiencing distant disease (for the last year of life for all women who die of breast cancer) (see Supplement Table 1).(33,40,41)

### Use and Harms

Use of services focused on the number of mammograms required for the screening strategy. Harms included false-positive mammograms, benign biopsies, and overdiagnosis. False-positive mammogram rates were calculated as mammograms read as abnormal or needing further work-up in women without cancer divided by the total number of screening mammograms. Benign biopsies were defined as biopsies among women with false-positive screening results; we assume 100% compliance with biopsy recommendations.(42) Overdiagnosis was defined as all cases that would not have been clinically detected in the absence of screening because of lack of progressive potential or death from competing non-breast cancer mortality. The impact of overdiagnosis on QALYs was captured by the disutility of being treated for cancer but dying of other causes.

### Statistical Analysis

For each model, strategies were ranked by the number of mammograms performed. We report the median use, benefits, and harms and range across models. We also obtained an efficiency frontier by plotting the sequence of points that represent the largest incremental percent breast cancer mortality reduction (or LYs or QALYs) per mammogram performed or harm entailed. Screening strategies that fell on this frontier were considered the most efficient (i.e., have the steepest slope such that no alternative exists that provides more benefit with less use/fewer harms).

Three models (E, GE, and W) also evaluated results based on combinations of breast cancer risk and density. Risk levels included: 1.3 (e.g., nulliparity or age at first live birth >30); (18,43) 2.0 (e.g., family history of one first degree relative); (18) or 4.0 times higher than average-risk (e.g., 2 or more first degree relatives).(18,44) Greater risk levels, such as seen with BRCA 1/2 mutations, were not considered since such groups have specific screening guidelines. We made the simplifying assumption that risk affected incidence, but not other aspects of disease.

Breast density was modeled as entirely fatty ("a"), scattered density ("b"), heterogeneously dense ("c") and extremely dense ("d"). Based on observed age-specific prevalence rates, density was assigned at age 40, and remained the same or decreased by one level at age 50 and again at age 65.(45) Density modified mammography sensitivity and specificity based on age, interval, and first vs. subsequent screening.(33) Density also modified the age-group specific (40-49, 50-64, and 65+) risk of developing breast cancer compared to average population density in the age-group (BCSC unpublished data).(44,46) Density was assumed to not affect molecular subtype or disease natural history. Density results were grouped into low ("a and b") and high density ("c and d") for presentation. The risk- and density-specific results were also compared to those for screening average-risk and density groups biennially from 50-74, since many guideline groups accept the latter.

In other analyses, two models (model E and GE) examined the impact of comorbidity on screening cessation using comorbidity-specific life expectancy. Examples of conditions that placed women in severe and moderate comorbidity groups included congestive heart failure and diabetes, respectively; the specific conditions and their associated life expectancies have been previously reported.(20,35,47) We compared results for continuing to screen biennially past age 74 among women with no or low comorbidity or stopping earlier than 74 for those with moderate or high comorbidity. These analyses included women who survived and were breast cancer-free up until the point where screening was to be extended or stopped.

Four models evaluated whether high disutility values would eliminate screening benefits. Finally, we evaluated the ability of the models to independently predict external trends and results (Supplement Figure 1 and Supplement Table 2).

*Role of the Funding Source*

We worked with US Preventive Services Task Force and Agency for Healthcare Research and Quality to develop the research questions. NCI investigators (KC, EF) collaborated in their role as scientific project officers. The agencies had no role in the study conduct or decision to submit the manuscript for publication.

## RESULTS

### Benefits in the Average-risk Population

The models produced consistent rankings of the screening strategies (Table 1). For instance, biennial screening from ages 50 to 74 yielded a median 25.8% reduction in breast cancer mortality compared to no screening (range: 24.1%-31.8). Annual screening led to slightly greater reductions in mortality than biennial strategies. However, biennial strategies maintained a median of 79.8%-81.3% of the breast cancer-specific mortality reduction of annual screening (range 68.3-98.9%) (Supplement Table 3).

Biennial screening also maintained the majority of annual benefits for LYS and QALYs and quality-adjustment did not change the ranking of strategies. Across all strategies, the largest decrement from quality-adjustment to life years was related to declines in general health as women aged; smaller decrements occurred due to the disutility of undergoing diagnostic evaluation of an abnormal screening exam and for having cancer. The disutility associated with screening itself had minimal impact on QALYs. (see 33)

The incremental benefits of initiating screening at age 40 were slightly greater than starting at age 50 in terms of breast cancer deaths averted with both annual and biennial screening (median 1.3 [range: 1.1-1.7] and 1.0 [0.8-1.7] per 1000 women screened, respectively) (Table 3). Initiating screening at age 45 yielded benefits intermediate be-

**Table 1** Ranking of Benefits (Percent Breast Cancer Mortality Reduction, LYs, QALYs) by Model and Screening Strategy Per 1000 Women Screened

| Strategies | Results per 1000 Women Screened | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | # of screens* | Percent breast cancer mortality reduction (vs. no screening) by model[1] | | | | | | Median (range across models) |
| | | D | E | G-E | M | S | W | |
| B 50-74 | 11,127 | 25.6% | 26.0% | 31.8% | 26.8% | 24.1% | 25.4% | 25.8% (24.1-31.8) |
| B 45-74 | 13,212 | 26.6% | 27.6% | 33.9% | 28.4% | 25.9% | 26.7% | 27.2% (25.9-33.9) |
| H 45-74 | 15,966 | 27.7% | 29.7% | 35.9% | 29.2% | 27.3% | 30.1% | 29.5% (27.3-35.9) |
| B 40-74 | 16,013 | 28.3% | 30.3% | 35.9% | 31.9% | 28.2% | 30.5% | 30.4% (28.2-35.9) |
| H 40-74 | 20,884 | 29.0% | 32.3% | 37.9% | 31.7% | 29.3% | 32.8% | 32.0% (29.0-37.9) |
| A 50-74 | 21,318 | 32.1% | 33.9% | 37.6% | 27.1% | 29.1% | 35.3% | 33.0% (27.1-37.6) |
| A 45-74 | 26,136 | 34.2% | 37.6% | 41.6% | 29.4% | 32.3% | 39.1% | 35.9% (29.4-41.6) |
| A 40-74 | 31,038 | 35.5% | 40.1% | 43.6% | 32.5% | 34.4% | 42.6% | 37.8% (32.5-43.6) |

[1]Without screening, the median probability of dying of breast cancer is 2.50% (range 1.50-3.20%). Thus, if a particular screening strategy leads to a 30% reduction in breast cancer mortality, this means that the probability of breast cancer mortality was reduced from 2.50% to 1.75%. This translates into 7.5 deaths averted per 1000 women screened. The absolute reduction in breast cancer deaths (i.e., deaths averted) vs. no screening for each strategy is included in Table 2.

| Strategies | Results per 1000 Women Screened | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | # of screens* | Years of Life Gained (vs. no screening) by model | | | | | | Median (range across models) |
| | | D | E | G-E | M | S | W | |
| B 50-74 | 11,127 | 153.8 | 94.0 | 140.5 | 146.5 | 104.2 | 74.6 | 122.4 (74.6-153.8) |
| B 45-74 | 13,212 | 168.4 | 107.7 | 161.2 | 171.3 | 115.2 | 84.0 | 138.2 (84.0-171.3) |
| H 45-74 | 15,966 | 175.3 | 117.9 | 170.2 | 171.4 | 125.1 | 95.7 | 147.7 (95.7-175.3) |
| B 40-74 | 16,013 | 183.7 | 123.7 | 172.4 | 194.8 | 131.6 | 98.8 | 152.0 (98.8-194.8) |
| H 40-74 | 20,884 | 191.1 | 137.6 | 187.2 | 211.5 | 141.0 | 110.9 | 164.1 (110.9-211.5) |
| A 50-74 | 21,318 | 180.0 | 125.9 | 167.3 | 156.3 | 133.3 | 104.3 | 144.8 (104.3-180.0) |
| A 45-74 | 26,136 | 201.3 | 149.3 | 196.7 | 177.8 | 154.2 | 123.0 | 166.0 (123.0-201.3) |
| A 40-74 | 31,038 | 217.1 | 168.8 | 213.5 | 218.1 | 170.1 | 140.5 | 191.8 (140.5-218.1) |

| Strategies | Results per 1000 Women Screened | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | # of screens* | QALYs Gained (vs. no screening) by model | | | | | | Median (range across models) |
| | | D | E | G-E | M | S | W | |
| B 50-74 | 11,127 | 114.5 | 67.3 | 100.1 | 109.6 | 71.9 | 47.1 | 86.0 (47.1-114.5) |
| B 45-74 | 13,212 | 123.8 | 75.6 | 114.4 | 129.4 | 78.8 | 51.9 | 96.6 (51.9-129.4) |
| H 45-74 | 15,966 | 126.6 | 80.9 | 118.3 | 128.5 | 84.5 | 58.3 | 101.4 (58.3-128.5) |
| B 40-74 | 16,013 | 133.7 | 85.4 | 120.1 | 148.1 | 89.1 | 60.4 | 104.6 (60.4-148.1) |
| H 40-74 | 20,884 | 134.2 | 91.0 | 126.1 | 159.4 | 92.5 | 64.8 | 109.3 (64.8-159.4) |
| A 50-74 | 21,318 | 127.0 | 84.1 | 111.4 | 113.2 | 87.5 | 62.4 | 99.5 (62.4-127.0) |
| A 45-74 | 26,136 | 138.9 | 97.3 | 129.5 | 129.4 | 99.5 | 71.7 | 114.5 (71.7-138.9) |
| A 40-74 | 31,038 | 146.6 | 107.3 | 137.2 | 160.6 | 107.6 | 80.0 | 122.4 (80.0-160.6) |

A=Annual B=Biennial H=Hybrid
*Strategies are ranked from the least to the most mammograms, where the number of mammograms is the median across models. Not all possible mammograms in the age interval are obtained since some women die from other causes before screening would occur.
†Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G-E (Georgetown U. –Einstein COM.), M (M.D. Anderson Cancer Center), S (Stanford U.), W (University of Wisconsin/Harvard)
‡Grey shaded areas in the table show strategies that are inferior or inefficient ("dominated") within a specific model; a strategy is classified as inferior or inefficient if there is another strategy that results in an equal or higher benefit (either percent mortality decline; LYG; or QALYs) with fewer harms (e.g., average screening exams).
§QALYs are adjusted for general health, diagnosis, screening and treatment.
||100% of women receive adjuvant systemic therapy based on recommended stage, ER/HER2-specific adjuvant therapy for pre- and post-menopausal women.

tween beginning at 40 and 50, although there were slightly greater incremental benefits when starting at age 45 (vs. 50) than starting at age 40 (vs. 45) (e.g., 10.6 vs. 8.0 and 15.4 vs. 7.9 QALYs for biennial and annual strategies, respectively) (Table 1).

### Harms in the Average-risk Population

All models projected more false-positive results, benign biopsies, and overdiagnosed cases under annual vs. biennial schedules and starting earlier than age 50 (Table 2). For instance, if biennial screening began at age 40 instead of age 50, for every 1000 women screened there would be a median of 1 more death averted, but 576 more false-positive results, 58 benign biopsies, and 2 additional overdiagnosed cases. Compared to screening initiation at age 45, starting screening at age 40 had 1 or fewer added deaths averted depending on interval, but more incremental harms.

### Efficiency Frontiers for Average-risk Populations

Efficiency frontier plots were used to graphically depict the balance between the number of mammograms and benefits (life years gained) of screening strategies. Biennial strategies starting at either age 40, 45, and 50 were all efficient (Figure 1, Supplemental Figure 2). Points that were close to, but fell below the frontier were less efficient than those on the frontier line. For example, compared to the point on the efficient frontier for biennial screening at age 45, the hybrid strategy of annual screening at 45 was less efficient than biennial screening starting at 40. This is because the hybrid strategy at 45 would require 405.8 more mammograms to gain an additional life year for every 1000 women screened compared to biennial screening at 45, while biennial screening starting at 40 only requires 189.5 extra mammograms to gain an additional life year.

Finally, annual screening from ages 50 to 74 was consistently inferior to other strategies (i.e., was inefficient, or dominated) since it yielded the same or fewer benefits than the next least intensive strategy depending on the measure of benefits, but required

**Table 2** Lifetime Benefits and Harms of Screening Strategies based on Starting Ages and Screening Intervals

| Strategy | Median number (range across models) per 1000 women screened (vs. no screening)* | | | | | |
|---|---|---|---|---|---|---|
| | Screens | Breast cancer deaths averted | False-positive screens | Benign breast biopsies | Over-diagnosed cases (invasive and DCIS) † ‡ | Percent of all cases over-diagnosed † ‡ |
| **Biennial** | | | | | | |
| 50-74 | 11,127 | 7 (4-9) | 953 (830-1325) | 146 (120-205) | 19 (11-34) | 12% (8–22) |
| 45-74 | 13,212 | 8 (4-9) | 1220 (930-1599) | 168 (120-221) | 19 (11-34) | 12% (8–22) |
| 40-74 | 16,013 | 8 (5-10) | 1529 (1100-1976) | 204 (140-264) | 21 (12-38) | 13% (9–24) |
| **Hybrid** | | | | | | |
| 45-74 | 15,966 | 8 (5-9) | 1520 (1160-1968) | 190 (140-250) | 21 (12-40) | 13% (8–25) |
| 40-74 | 20,884 | 9 (5-10) | 2106 (1480-2623) | 245 (170-309) | 23 (12-44) | 14% (9–27) |
| **Annual** | | | | | | |
| 50-74 | 21,318 | 9 (5-10) | 1798 (1706-2445) | 228 (219-317) | 25 (12-68) | 15% (8–36) |
| 45-74 | 26,136 | 9 (6-11) | 2355 (2185-3087) | 247 (230-329) | 28 (12-74) | 17% (9–38) |
| 40-74 | 31,038 | 10 (6-11) | 2941 (2550-3742) | 303 (260-388) | 30 (13-77) | 18% (9–39) |

*In all scenarios, 100% of women receive adjuvant systemic therapy based on recommended stage, ER/HER2-specific adjuvant therapy for pre- and post-menopausal women.

†Over-diagnosed cases are defined as cases that would not have been clinically detected in the absence of screening

(i.e., cases that do not die from breast cancer because of lack of progressive potential or death from competing non-breast cancer mortality). The result includes DCIS and invasive overdiagnosis. Over-diagnosis is calculated by comparing cases detected in the screening scenario to those detected in the non-screened scenario. Model S is excluded since it does not include DCIS. The percent overdiagnosis is calculated as the percent of all cases detected in the screening strategy that are overdiagnosis.

‡The upper range for all over diagnosis estimates is based on model M results. Model M generates very high overdiagnosis based on the assumption that incidence in the absence of screening has essentially remained flat since 1975-79, with virtually all of the increases over time attributable to screening. The other models use some form of an age-period-cohort model for incidence in the absence of screening, where some of the increases in incidence are due to screening and some to changes in risk factors (e.g., use of hormone replacement therapy), generating lower rates of overdiagnosis. Other sources of variation across models are related to assumptions about the proportions of DCIS cases that never progress to invasive cancer or the number of early invasive cancers that might be nonprogressive. Generally, models that assume higher proportions of DCIS and/or invasive cancer to be nonprogressive generate higher estimates of overdiagnosis than models that assume less nonprogressive disease. Unfortunately, the underlying incidence in the absence of screening and the proportion and types of tumors that are nonprogressive are unknown and unobservable. Therefore, the different results across models based on their respective assumptions provide a range of possible overdiagnosis.

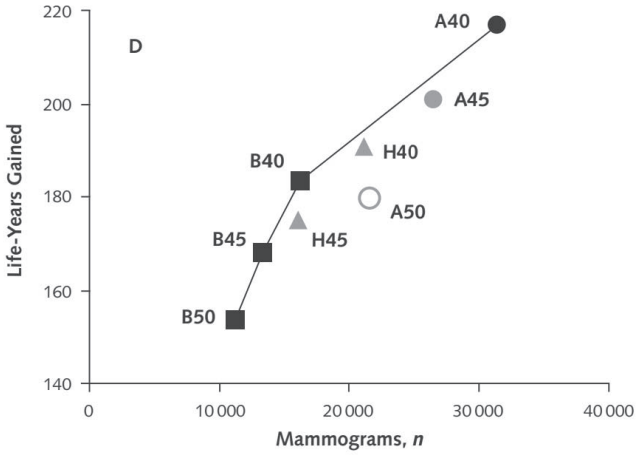



**Figure 1** Efficiency frontier for life-years gained versus mammograms performed per 1000 women in model D (Dana-Farber Cancer Institute).
Legend for Figure 1. Efficiency Frontier
Efficiency frontier graphs for all models are shown in Appendix Figure 2 (available at ww.annals.org). This graph plots the average gain in life-years per additional mammogram performed per 1000 women for each screening strategy (vs. no screening) in model D. Biennial strategies are indicated with a square; hybrid strategies (annual in the 40s followed by biennial from 50 to 74 years of age) with a triangle; and annual strategies with a circle. Efficient strategies were plotted (i.e., those in which increases in mammography use resulted in greater life-years gained than the next less intensive strategy). The line represents the "efficiency frontier" by joining efficient strategies in which increases in mammography use resulted in greater life-years gained than the next less intensive efficient strategy. Strategies on this line would be considered efficient because they achieve the greatest gain in benefit (life-years gained) per harm or use of mammograms. Strategies that use more mammograms but still have small benefits (i.e., a shallower slope than the next best strategy) are considered to be less efficient (i.e., weakly dominated). When and if the slope in the efficiency frontier plot levels off, it means that the additional life-years gained per increase in mammography are small relative to the previous strategies and could indicate a point at which additional screening might be considered as having a low return (or additional benefit). There is no definitive inflection point across the models for the strategies or metrics evaluated. Black strategies are efficient; gray strategies close to the efficiency frontier are less efficient; and open gray strategies are inefficient (inferior, or dominated). Reference (33) provides efficiency frontiers for other harm and benefit metrics.

more mammograms or entailed more harms. These above patterns were generally seen with other harm and benefit metrics (see Supplement Figure 2).

## Sensitivity Analyses for Average-Risk Populations

Varying the disutilities for usual health, screening, diagnosis, and treatment did not affect strategy rankings for average-risk populations and QALY gains persisted under all screening strategies, although their magnitude decreased.

**Table 3** Incremental Changes in Breast Cancer Deaths Averted by Interval, Age of Screening Initiation, and Model

| | Annual | | Biennial | |
|---|---|---|---|---|
| | Number of breast cancer deaths averted/1000 women (% breast cancer mortality reduction) | | | |
| **Model** | Start at 40 vs. 50 | Start at 45 vs. 50 | Start at 40 vs. 50 | Start at 45 vs. 50 |
| **D** | 1.1 (3.4%) | 0.6 (2.1%) | 0.9 (2.7%) | 0.3 (1.0%) |
| **E** | 1.5 (6.2%) | 0.9 (3.6%) | 1.0 (4.3%) | 0.4 (1.6%) |
| **G-E** | 1.5 (6.0%) | 1.0 (4.0%) | 1.0 (4.1%) | 0.5 (2.2%) |
| **M** | 1.7 (5.3%) | 0.7 (2.3%) | 1.7 (5.1%) | 0.5 (1.6%) |
| **S** | 1.1 (5.2%) | 0.7 (3.1%) | 0.9 (4.1%) | 0.4 (1.7%) |
| **W** | 1.1 (7.3%) | 0.6 (3.8%) | 0.8 (5.1%) | 0.2 (1.3%) |
| **Median** | 1.3 (5.7%) | 0.7 (3.4%) | 1.0 (4.2%) | 0.4 (1.6%) |

*Incremental difference between starting at age 40 or 45 vs. 50. Annual is comparing A40-74 (or 45-74) to A50-74; biennial is comparing B40-74 (or 45-74) to B50-74. Hybrid strategies are compared to B50-74, therefore for those incremental comparisons the hybrid results are the same as the annual results

## Harms and Benefits by Risk Level

The balance of harms and benefits differed by risk group, with women who had higher-risk having lower rates of false-positives and higher gains from screening than lower-risk groups. Screening higher-risk women also yielded a lower proportion of overdiagnosed cases per breast cancer death averted than screening average-risk women. However, annual screening from ages 50 to 74 had the same or less benefit and more harms than other strategies at all risk levels.(33)

For women with a 2- to 4-fold increase in risk, annual screening starting at age 40 or 45 had similar or more favorable harm-to-benefit ratios (based on false-positives) as biennial screening of average-risk women from 50-74. For instance, for every 1000 average-risk women screened biennially from 50-74, there would be 226.5 (range: 169.9-267.0) false-positives per death averted. If women with a two-fold increase in risk began annual screening at age 40, their corresponding ratio would be slightly more favorable at 200.7 (range: 177.5-232.2). For women with a 1.3-fold increase in risk, biennial screening starting at age 40 had similar harm-to-benefit ratios as biennial screening of average-risk women from ages 50-74.

## Benefits and Harms by Breast Density Group

Breast density (low vs. high) changed absolute benefits, but annual screening from 50-74 remained inefficient across breast density groups. Women in the low-density group had a greater proportion of their cancers detected due to greater digital mammography sensitivity, and therefore a greater breast cancer-specific mortality reduction than the high-density group. However, women in the high-density group had a greater absolute

number of cancers detected because their risk of cancer was higher, leading to more life years saved among women in the high-density than the low-density group (33)).

### Benefits and Harms by Comorbidity

For women with no comorbidity, biennial screening could continue to age 78 or 80 and still have similar harm-to-benefit ratios as screening women with average non-breast cancer mortality biennially from 50-74. However, for women with moderate to severe comorbidity, the comparable ratios were equivalent at about age 68 (33).

## DISCUSSION

This study used six established models to estimate the potential efficacy of different US breast cancer screening strategies. All six models demonstrated that screening initiation at age 40 has some benefits for average-risk populations, but also higher levels of harms than strategies starting at age 50. The findings also suggest that comorbidity levels could be used to tailor the age of screening cessation. Biennial screening strategies were the most efficient, but annual screening could be considered from ages 40-74 in groups with a two to four-fold higher than average-risk.

Results from all models indicated that digital mammography screening of average-risk women in their 40's modestly lowers breast cancer-specific mortality and extends the length and quality of life, even after considering disutilities related to the screening process. The absolute benefits of starting screening in the 40's varied somewhat based on model structure and assumptions, but were consistent with observations from randomized trials.(6) However, starting at age 40 vs. 45 was associated with increasing incremental harms relative to the increase in benefits. Thus, decisions about initiating screening before age 50 may depend on the weight attached to screening benefits and harms.

Consistent with other analyses of screening upper age limits,(20,48-50) and other recommendations,(12,51) our results suggested that the balance of harms and benefits of screening was affected by competing non-breast cancer mortality, so that age of screening cessation could be tailored by comorbidity levels.

Similar to our 2009 analysis,(1) biennial strategies are most consistently efficient. Screening annually from ages 50-74 had the same or fewer benefits for any given harm for all population groups in virtually all models, and would be considered inefficient. However, annual screening in the 40's followed by biennial screening at age 50, or the most intensive schedule evaluated (annual screening 40-74) were also efficient or close to being efficient. Additionally, annual screening of women with a two- to four-fold increased risk (e.g., due to non-BRCA related family history) from ages 40-74 had com-

parable harm-to-benefit ratios as did biennial screening from age 50 to 74 in average-risk populations.

The results also suggest that benefits of screening vary by breast density, at least when grouped into low/high categories. Women with dense breasts have a higher risk of developing cancer and absolute detection rate, but lower relative detection. (19,52) This is because digital sensitivity, while optimized for density, is still lower in women with dense than non-dense breasts.(53-56) Improving outcomes for women with dense breasts (55) may require new innovations in imaging (57-60) or identification of risk biomarkers.(61,62)

This analysis extends our prior work by explicitly considering overdiagnosis as a screening harm. Depending on screening strategy, the models estimated that 2-12% of invasive and 30% to 50% of DCIS cases might represent overdiagnosis. While the models differed in absolute estimates, they agreed on how overdiagnosis affected the ranking of strategies and the finding that the majority of overdiagnosed cases were DCIS. The model results for overdiagnosis are not directly comparable to other published estimates (8,63) since the models followed women for their entire lives. The models also made assumptions about unobservable input parameters related to natural history. While there is no agreement on methods to estimate overdiagnosis (64) or on its true rate,(65,66) there is agreement that it is an important harm. Active surveillance for DCIS with a low risk of progression is one potential future approach to reduce harms from DCIS overdiagnosis. More information is also needed on consumer knowledge of and willingness to risk overdiagnosis.(67)

Overall, this study has several important strengths including collaboration of six long-established, independent modeling groups, use of well-calibrated models that reproduce temporal epidemiological trends and a screening trial result, inclusion of digital technology, incorporation of increasingly effective treatments, and consideration of quality of life, risk factors, breast density, and comorbidity.(68) The conclusions about the ranking of screening strategies are robust and should provide greater credibility than inferences based on one model alone.

Our study also had limitations. First, to evaluate program efficacy we assumed 100% adherence to screening, prompt evaluation of abnormal results, and full use of optimal treatment. Actual benefits will fall short of our projected results since adherence is not perfect. Second, we only focused on hybrid strategies for women in their 40's. Alternative hybrid strategies may be important to examine in future research. Third, the analysis also did not consider other imaging technologies for average-risk populations or for groups with high breast density, such as ultrasound, (69) computer-aided detection,(70) tomosynthesis, or magnetic resonance imaging (MRI). Data on tomosynthesis performance and needs for radiologist re-training are still emerging.(58) Fourth, we did not model any radiation-induced breast cancers due to more intensive mammography

schedules.(71) Fifth, we assumed that risk factors influenced the incidence of disease, but not its natural history. Sixth, certain risk factors, such as family history, are age-dependent in their effects.(18,72) Since we held relative risk levels constant over age, our benefit estimates could be over- or under-estimated for specific risk factors.(17) Seventh, we did not consider polygenic risk,(73,74) or explicitly model menopausal status; we used age 50 as a proxy for the average age of menopause. Additionally, the analysis did not include screening program costs or utility estimates specific to some of the newest treatments. Finally, compared to our earlier research,(1) the models all estimated similar, but somewhat greater breast cancer-specific mortality reductions (for example, a median 22% vs. 25.8% reduction with biennial screening from 50-74 in 2009 vs. current models, respectively). The primary reasons for this modeled improvement relate to the increased sensitivity of digital vs. film mammography, advances in molecular-targeted therapies, and changes in underlying breast cancer trends.

Overall, the six models conclude that biennial screening strategies are the most efficient. Choices about optimal ages of initiation (and cessation) and screening intervals will ultimately depend on program goals, the weight attached by the decision-maker to screening harms and benefits,(75) and considerations of efficiency.

**Acknowledgments**

The authors thank CISNET consultants Elizabeth Burnside, Allison Kurian, Donald Weaver, and Diana Buist for review of earlier versions of the models for structure, parameters, and assumptions for clinical face validity; Jennifer Croswell from the Agency for Healthcare Research and Quality for assistance with analysis and review of results; members of the U.S. Preventive Services Task Force and the Oregon Evidence-based Practice Center for comments on earlier versions of this research; and Jessica Garshell for data processing. The authors thank Adrienne Ryans for manuscript preparation.

The authors thank John Wong, Stuart Baker, William Lawrence, and Tom Trikalinos for helpful suggestions on earlier versions of this paper.

**POTENTIAL CONFLICTS OF INTEREST: NONE DISCLOSED**

## REFERENCES


(1) Mandelblatt J, Cronin K, Bailey S, Berry DA, de Koning H, Draisma G et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. Ann Inten Med 2009; 151(10):738-747.

(2) Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2013; 6:CD001877.

(3) Biller-Andorno N, Juni P. Abolishing mammography screening programs? A view from the Swiss Medical Board. N Engl J Med 2014; 370(21):1965-1967.

(4) Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 2002; 359(9310):909-919.

(5) Tabar L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF et al. The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. Radiol Clin North Am 2000; 38(4):625-651.

(6) Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. Lancet 2006; 368(9552):2053-2060.

(7) Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M et al. Effect of screening and adjuvant therapy on mortality from breast cancer. New Engl J Med 2005; 353(17):1784-1792.

(8) Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. BMJ 2014; 348:g366.

(9) Paci E, Broeders M, Hofvind S, Puliti D, Duffy SW. European breast cancer service screening outcomes: a first balance sheet of the benefits and harms. Cancer Epidemiol Biomarkers Prev 2014; 23(7):1159-1163.

(10) Smith RA. The value of modern mammography screening in the control of breast cancer: understanding the underpinnings of the current debates. Cancer Epidemiol Biomarkers Prev 2014; 23(7):1139-1146.

(11) Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med 2015; 372(24):2353-2358.

(12) Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009; 151(10):716-236.

(13) Broeders M, Moss S, Nystrom L, Njor S, Jonsson H, Paap E et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. J Med Screen 2012; 19 Suppl 1:14-25.

(14) Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. Lancet Oncol 2015. Sep;16(9):1123-32.

(15) US Food and Drug Administration. Available from: http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/default.htm. June 2013. Silver Spring, Maryland. US Department of Health and Human Services. Accessed January 2015.

(16) National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice

Guidelines in Oncology - Breast Cancer. Available from: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp. 2014. Accessed January 2015.

(17) van Ravesteyn NT, Miglioretti DL, Stout NK, Lee SJ, Schechter CB, Buist DS et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: a comparative modeling study of risk. Ann Intern Med 2012; 156(9):609-617.

(18) Nelson HD, Zakher B, Cantor A, Fu R, Griffin J, O'Meara ES et al. Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. Ann Intern Med 2012; 156(9):635-648.

(19) Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. Ann Intern Med 2011; 155(1):10-20.

(20) 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 2014; 161(2):104-112.

(21) de Gelder R, Heijnsdijk EA, Fracheboud J, Draisma G, de Koning HJ. The effects of population-based mammography screening starting between age 40 and 50 in the presence of adjuvant systemic therapy. Int J Cancer 2015; 137(1):165-172.

(22) Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR. Assessing the effectiveness of health interventions for cost-effectiveness analysis. Panel on Cost-Effectiveness in Health and Medicine. J Gen Intern Med 1997; 12(9):551-558.

(23) Munoz D, Near AM, van Ravesteyn NT, Lee SJ, Schechter CB, Alagoz O et al. Effects of screening and systemic adjuvant therapy on ER-specific US breast cancer mortality. J Natl Cancer Inst 2014 Sep 24;106(11).

(24) Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL. The Wisconsin Breast Cancer Epidemiology Simulation Model. J Natl Cancer Inst Monogr 2006;(36):37-47.

(25) Mandelblatt J, Schechter CB, Lawrence W, Yi B, Cullen J. The SPECTRUM population model of the impact of screening and treatment on U.S. breast cancer trends from 1975 to 2000: principles and practice of the model methods. J Natl Cancer Inst Monogr 2006; 36:47-55.

(26) Berry DA, Inoue L, Shen Y, Venier J, Cohen D, Bondy M et al. Modeling the impact of treatment and screening on U.S. breast cancer mortality: a Bayesian approach. J Natl Cancer Inst Monogr 2006;(36):30-36.

(27) Plevritis SK, Sigal BM, Salzman P, Rosenberg J, Glynn P. A stochastic simulation model of U.S. breast cancer mortality trends from 1975 to 2000. J Natl Cancer Inst Monogr 2006;(36):86-95.

(28) Lee S, Zelen M. A stochastic model for predicting the mortality of breast cancer. J Natl Cancer Inst Monogr 2006;(36):79-86.

(29) Tan SY, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The MISCAN-Fadia continuous tumor growth model for breast cancer. J Natl Cancer Inst Monogr 2006;(36):56-65.

(30) Clarke LD, Plevritis SK, Boer R, Cronin KA, Feuer EJ. A comparative review of CISNET breast models used to analyze U.S. breast cancer incidence and mortality trends. J Natl Cancer Inst Monogr 2006;(36):96-105.

(31) Gangnon RE, Sprague BL, Stout NK, Alagoz O, Weedon-Fekjaer H, Holford TR et al. The Contribution of Mammography Screening to Breast

Cancer Incidence Trends in the United States: An Updated Age-period-cohort Model. Can Epi Biomarkers Prev 2015. Jun;24(6):905-12

(32) Peto R, Davies C, Godwin J, Gray R, Pan HC, Clarke M et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Lancet 2012; 379(9814):432-444.

(33) Mandelblatt JS, Cronin K, de Koning H, Miglioretti DL, Schechter CS, Stout N. Modleing Technical Report. Collaborative modeling of U.S. breast cancer screening strategies. AHRQ Publication No. 14-05201-EF-4. April 2015 Available from: http://www.uspreventiveservicestaskforce.org/Page/Document/modeling-report-collaborative-modeling-of-us-breast-cancer-1/breast-cancer-screening1. Accessed May 2015.

(34) Chang Y, Schechter CB, van Ravesteyn NT, Near AM, Heijnsdijk EA, Adams-Campbell L et al. Collaborative modeling of the impact of obesity on race-specific breast cancer incidence and mortality. Breast Cancer Res Treat 2012; 136:823-835.

(35) Cho H, Klabunde CN, Yabroff KR, Wang Z, Meekins A, Lansdorp-Vogelaar I et al. Comorbidity-adjusted life expectancy: a new tool to inform recommendations for optimal screening strategies. Ann Intern Med 2013; 159(10):667-676.

(36) Holford TR, Cronin KA, Mariotto AB, Feuer EJ. Changing patterns in breast cancer incidence trends. J Natl Cancer Inst Monogr 2006;(36):19-25.

(37) Kerlikowske K, Zhu W, Hubbard RA, Geller B, Dittus K, Braithwaite D et al. Outcomes of screening mammography by frequency, breast density, and post-menopausal hormone therapy. JAMA Intern Med 2013; 173(9):807-816.

(38) Dittus K, Geller B, Weaver DL, Kerlikowske K, Zhu W, Hubbard R et al. Impact of mammography screening interval on breast cancer diagnosis by menopausal status and BMI. J Gen Intern Med 2013; 28(11):1454-1462.

(39) Hanmer J, Lawrence WF, Anderson JP, Kaplan RM, Fryback DG. Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores. Med Decis Making 2006; 26(4):391-400.

(40) Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective cost-effectiveness analysis of screening mammography. J Natl Cancer Inst 2006; 98(11):774-782.

(41) de Haes JC, de Koning HJ, Van Oortmarssen GJ, van Agt HM, De Bruyn AE, van der Maas PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. Int J Cancer 1991; 49:538-544.

(42) Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM et al. Performance benchmarks for screening mammography. Radiology 2006; 241(1):55-66.

(43) Kerlikowske K, Walker R, Miglioretti DL, Desai A, Ballard-Barbash R, Buist DS. Obesity, mammography use and accuracy, and advanced breast cancer risk. J Natl Cancer Inst 2008; 100(23):1724-1733.

(44) Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst 2006; 98(17):1204-1214.

(45) Sprague BL, Gangnon RE, Burt V, Trentham-Dietz A, Hampton JM, Wellman RD et al. Prevalence of mammographically dense breasts in the United States. J Natl Cancer Inst 2014; 106(10).

(46) Tice JA, O'Meara ES, Weaver DL, Vachon C, Ballard-Barbash R, Kerlikowske K. Benign breast disease, mammographic breast density, and the risk of breast cancer. J Natl Cancer Inst 2013; 105(14):1043-1049.

(47) Mariotto AB, Wang Z, Klabunde CN, Cho H, Das B, Feuer EJ. Life tables adjusted for comorbidity more accurately estimate noncancer survival for recently diagnosed cancer patients. J Clin Epidemiol 2013; 66(12):1376-1385.

(48) Walter LC, Covinsky KE. Cancer screening in elderly patients: a framework for individualized decision making. JAMA 2001; 285(21):2750-2756.

(49) Schonberg MA, Davis RB, McCarthy EP, Marcantonio ER. External validation of an index to predict up to 9-year mortality of community-dwelling adults aged 65 and older. J Am Geriatr Soc 2011; 59(8):1444-1451.

(50) van Ravesteyn N, 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. J Nat Cancer Inst 2015; 107(7).

(51) Smith RA, Cokkinides V, Brawley OW. Cancer screening in the United States, 2012: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin 2012; 62(2):129-142.

(52) Vilaprinyo E, Forne C, Carles M, Sala M, Pla R, Castells X et al. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. PLoS One 2014; 9(2):e86858.

(53) Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005; 353(17):1773-1783.

(54) Kerlikowske K, Hubbard RA, Miglioretti DL, Geller BM, Yankaskas BC, Lehman CD et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. Ann Intern Med 2011; 155(8):493-502.

(55) Kerlikowske K, Zhu W, Tosteson AN, Sprague BL, Tice JA, Lehman CD et al. Identifying women with dense breasts at high risk for interval cancer: a cohort study. Ann Intern Med 2015; 162(10):673-681.

(56) Braithwaite D, Zhu W, Hubbard RA, O'Meara ES, Miglioretti DL, Geller B et al. 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(5):334-341.

(57) Lee CI, Lehman CD. Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. J Am Coll Radiol 2013; 10(12):913-917.

(58) Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA 2014; 311(24):2499-2507.

(59) Park JY, Yi SY, Park HJ, Kim MS, Kwon HJ, Park NH et al. Breast-specific gamma imaging: correlations with mammographic and clinicopathologic characteristics of breast cancer. AJR Am J Roentgenol 2014; 203(1):223-228.

(60) Rechtman LR, Lenihan MJ, Lieberman JH, Teal CB, Torrente J, Rapelyea JA et al. Breast-specific gamma imaging for the detection of breast cancer in dense versus nondense breasts. AJR Am J Roentgenol 2014; 202(2):293-298.

(61) Matamala N, Vargas MT, Gonzalez-Campora R, Minambres R, Arias JI, Menendez P et al. Tumor microRNA expression

profiling identifies circulating microRNAs for early breast cancer detection. Clin Chem. 2015 Aug;61(8):1098-106.

(62) Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst 2015; 107(5).

(63) Welch HG, Passow HJ. Quantifying the benefits and harms of screening mammography. JAMA Intern Med 2014; 174(3):448-454.

(64) Etzioni R, Gulati R, Mallinger L, Mandelblatt J. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. Ann Intern Med 2013; 158(11):831-838.

(65) Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. BMJ 2015; 350:g7773.

(66) Etzioni R, Xia J, Hubbard R, Weiss NS, Gulati R. A reality check for overdiagnosis estimates associated with breast cancer screening. J Natl Cancer Inst 2014; 106(12).

(67) Moynihan R, Nickel B, Hersch J, Beller E, Doust J, Compton S et al. Public opinions about overdiagnosis: a national community survey. PLoS One 2015; 10(5):e0125165.

(68) Elmore JG, Harris RP. The harms and benefits of modern screening mammography. BMJ 2014; 348:g3824.

(69) Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Alagoz O et al. Benefits, Harms, and Cost-Effectiveness of Supplemental Ultrasonography Screening for Women With Dense Breasts. Ann Intern Med 2015 Feb 3;162(3):157-66

(70) Fenton JJ, Xing G, Elmore JG, Bang H, Chen SL, Lindfors KK et al. Short-term outcomes of screening mammography using computer-aided detection: a population-based study of medicare enrollees. Ann Intern Med 2013; 158(8):580-587.

(71) Miglioretti DL, Lange J, van Ravesteyn N, van den Broek JJ, Lee CI, Melnikow J et al. Radiation-Induced Breast Cancer and Breast Cancer Death From Mammography Screening [Abstract]. Rockville, MD: Agency for Healthcare Research and Quality. 2015.

(72) Trentham-Dietz A, Sprague BL, Hampton JM, Miglioretti DL, Nelson HD, Titus LJ et al. Modification of breast cancer risk according to age and menopausal status: a combined analysis of five population-based case-control studies. Breast Cancer Res Treat 2014; 145(1):165-175.

(73) Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet 2013; 45(4):392.

(74) Stevens KN, Vachon CM, Couch FJ. Genetic susceptibility to triple-negative breast cancer. Cancer Res 2013; 73(7):2025-2030.

(75) Hoffmann TC, Del MC. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA Intern Med 2015; 175(2):274-286.

## Appendix Model validation

Each model has a different structure and assumptions and some varying input variables, so no single method can be used to validate results against an external standard. Therefore, we used several approaches. First, considering actual screening and treatment patterns instead of the efficacy strategies simulated in the base case, we compared model projections of incidence, breast cancer–specific mortality, and stage distribution with those reported by the Surveillance, Epidemiology, and End Results program for 1975 to 2010. In our previous work, results of each model accurately projected trends for incidence and breast cancer–specific mortality by ER status for 1975 to 2000 (23). Next, we approximated the Age screening trial (6), assuming perfect adherence to invitations for annual screening with 13-year follow-up of women aged 40 to 49 years (6). Finally, we examined the consistency of results across models. Using inputs for actual dissemination of screening and treatment in the United States, the models captured the major trends in incidence and the general shape of breast cancer–specific mortality decreases over time (Appendix Figure 1). They also closely matched current stage distribution (not shown) and the Age trial results (Appendix Table 2) (6, 33). Thus, the models replicated patterns of observed US incidence and breast cancer–specific mortality over time. The models also estimated similar breast cancer–specific mortality reduction as that observed among women aged 40 to 49 years who actually attended screening in the Age trial, although the model results are slightly more optimistic than the trial because the models assume 100% screening and use of the most effective systemic regimens (6). Overall, use of 6 models to project a range of plausible screening outcomes provides implicit cross-validation, with the range of results from the models as a measure of uncertainty.

*Appendix Table 1.* Utility Input Parameter Values

**Utilities for Cancer-Related States***

| State | Utility | Disutility (Worst Case 150%, 200%) | Duration | Unit |
|---|---|---|---|---|
| Cancer treatment for local or DCIS | 0.9 | 0.1 (0.15, 0.20) | 2 | Year |
| Cancer treatment for regional | 0.75 | 0.25 (0.375, 0.50) | 2 | Year |
| Cancer treatment for distant | 0.6 | 0.4 (0.6, 0.8) | Until death | - |
| Screening attendance (routine screening) | 0.994 | 0.006 (0.009, 0.012) | 1 | Week |
| Diagnostic phase (evaluation of positive screen) | 0.895 | 0.105 (0.158, 0.210) | 5 | Weeks |

**Age-Specific Utilities for General Health in U.S. Women†**

| Age | Healthy Base Value (Range) |
|---|---|
| 20 y | 0.913 (0.905-0.920) |
| 25 y | 0.913 (0.905-0.920) |
| 30 y | 0.893 (0.886-0.900) |
| 35 y | 0.893 (0.886-0.900) |
| 40 y | 0.863 (0.855-0.871) |
| 45 y | 0.863 (0.855-0.871) |
| 50 y | 0.837 (0.829-0.846) |
| 55 y | 0.837 (0.829-0.846) |
| 60 y | 0.811 (0.800-0.822) |
| 65 y | 0.811 (0.800-0.822) |
| 70 y | 0.771 (0.758-0.784) |
| 75 y | 0.771 (0.758-0.784) |
| 80 y | 0.724 (0.701-0.747) |
| 85 y | 0.724 (0.701-0.747) |

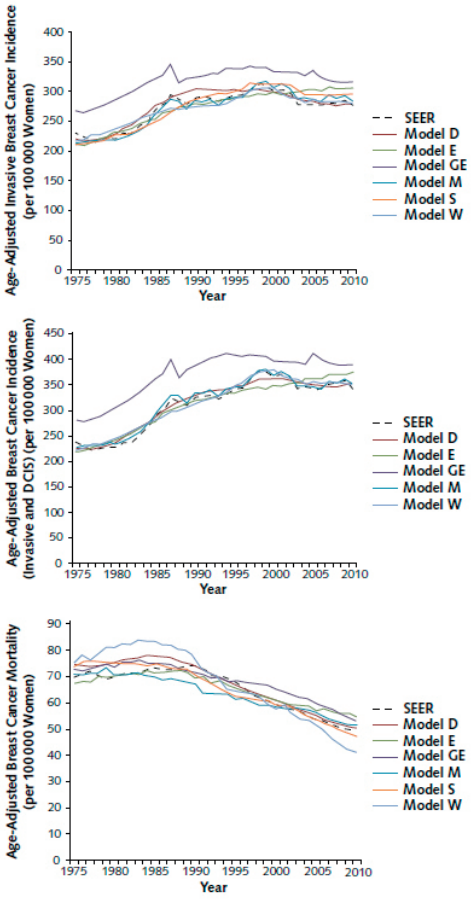
DCIS = ductal carcinoma in situ.
* From references 40 and 41.
† Values from the EuroQoL-5D quality-of-life questionnaire (39).

*Appendix Table 2.* Approximation of the Age Trial With 13-y Follow-up, by Model*

| Model | Relative Risk for Breast Cancer Death With 100% Screening† |
|---|---|
| D | 0.75 |
| E | 0.73 |
| GE | 0.65 |
| M | 0.72 |
| S | 0.69 |
| W | 0.71 |
| Median (range) | 0.72 (0.65-0.75) |

D = Dana-Farber Cancer Institute; E = Erasmus Medical Center; GE = Georgetown University Medical Center and Albert Einstein College of Medicine; M = MD Anderson Cancer Center; S = Stanford University; W = University of Wisconsin and Harvard Medical School. * Projection of relative risk of breast cancer death with annual screening from age 40 to 49 y; biennial at age 50 and 52 y versus a control group with biennial screening at age 50 and 52 y. Because the models are estimating mortality reduction with actual screening, model estimates are most comparable to the Age trial results (6) among women who actually attended screening. Model results show more benefit than observed in the trial because the models assume that 100% of women complied with the trial-specified screening schedule. In reality, not all women who were invited attended screening, and among those who attended, many did not attend all scheduled screening rounds. In addition, the models assumed 100% receipt of the most effective treatments. † Age trial invitation results (intention to treat): relative risk, 0.83 (95% CI, 0.66–1.04). Age trial results for women who actually were screened: relative risk, 0.76 (CI, 0.51–1.01).

***Appendix Figure 1.*** Modeled versus observed incidence of breast cancer and breast cancer-specific mortality in women aged 40 to 100 years.



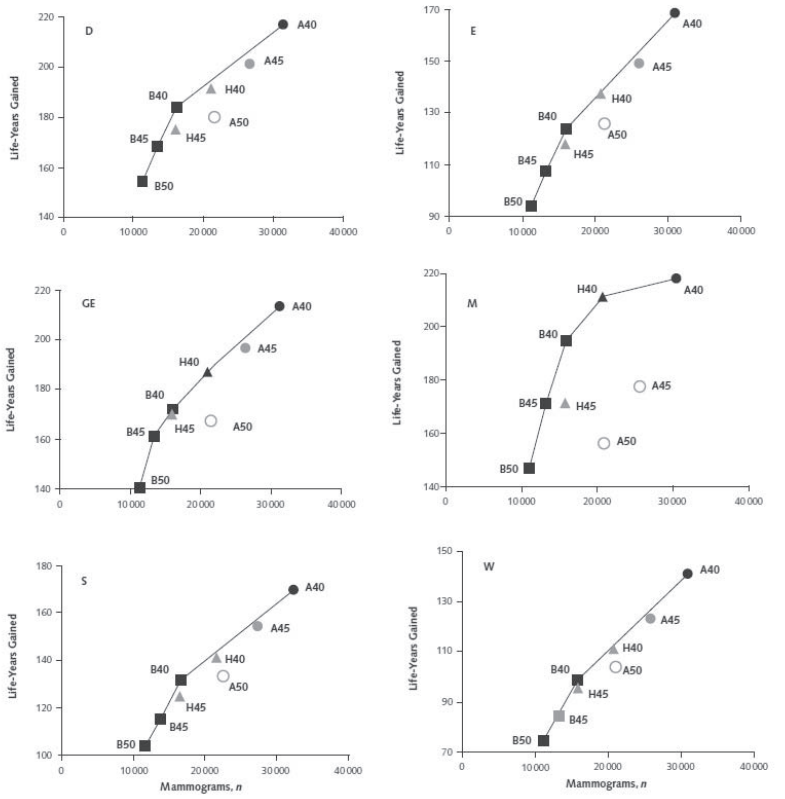

The models closely estimate observed U.S. trends in incidence of invasive disease (top), incidence of invasive disease and DCIS (middle)*, and breast cancer–specific mortality (bottom). Using inputs for actual dissemination of screening and treatment in the United States, the models all captured the major trends in incidence over time. Early increases with the advent of mammography in the mid-1980s are seen, followed by a downturn in the 2000s and then a leveling off. The models also captured the general shape of decreases in breast cancer–specific mortality over time. All models show an increase in incidence with the introduction of mammography screening. Model GE has a steep peak in incidence in 2005 owing to the specific method for capturing the transition from plain film to digital mammography, because digital mammography has higher sensitivity and detection of ductal carcinoma in situ than plain film mammography; other models include a more gradual transition surrounding this period. D = Dana- Farber Cancer Institute; DCIS = ductal carcinoma in situ; E = Erasmus Medical Center; GE = Georgetown University Medical Center and Albert Einstein College of Medicine; M = MD Anderson Cancer Center; QALY = quality-adjusted life-year; S = Stanford University; SEER = Surveillance, Epidemiology, and End Results; W = University of Wisconsin and Harvard Medical School. * Model S does not include DCIS.

*Appendix Table 3.* Annual Mortality Reduction Maintained by Biennial Screening, by Strategy and Model

| Age at Screening | Mortality Reduction, % | | | | | | |
|---|---|---|---|---|---|---|---|
| | Model D | Model E | Model GE | Model M* | Model S | Model W | Median |
| 50-74† | 79.8 | 76.7 | 84.6 | 98.9 | 82.8 | 72.0 | 81.3 |
| 45-74‡ | 77.8 | 73.4 | 81.5 | 96.6 | 80.2 | 68.3 | 79.0 |
| 40-74§ | 79.7 | 75.6 | 82.3 | 98.2 | 82.0 | 71.6 | 80.8 |

D = Dana-Farber Cancer Institute; E = Erasmus Medical Center; GE = Georgetown University Medical Center and Albert Einstein College of Medicine; M = MD Anderson Cancer Center; S = Stanford University; W = University of Wisconsin and Harvard Medical School. * Model M does not include a natural history component. On the basis of a combination of assumptions about underlying incidence trends in the absence of screening, it essentially yields a long lead time for invasive cancer; thus, all cancers found with annual screening can also be detected with biennial screening. † Percentage of reduction with annual screening in women aged 50-74 y that is maintained by biennial screening in women aged 50-74 y is calculated as the percent mortality reduction with biennial screening in women aged 50-74 y divided by the percent mortality reduction with annual screening in women aged 50-74 y. § Percentage of reduction with annual screening in women aged 45-74 y that is maintained by biennial screening in women aged 45-74 y is calculated as the percent mortality reduction with biennial screening in women aged 45-74 y divided by the percent mortality reduction with annual screening in women aged 45-74 y. § Percentage of reduction with annual screening in women aged 40-74 y that is maintained by biennial screening in women aged 40-74 y is calculated as the percent mortality reduction with biennial screening in women aged 40-74 y divided by the percent mortality reduction with annual screening in women aged 40-74 y.

*Appendix Figure 2.* Efficiency frontier for life-years gained versus mammograms performed for each screening strategy, by model.




The average gain in life-years per additional mammogram performed per 1000 women for each screening strategy (vs. no screening). Biennial strategies are indicated with a square; hybrid strategies (annual in the 40s followed by biennial from 50 to 74 years of age) with a triangle; and annual strategies with a circle. Efficient strategies were plotted (those in which increases in mammography use resulted in greater life-years gained than the next least-intensive strategy). The line represents the "efficiency frontier" by joining efficient strategies in which increases in mammography use resulted in greater life-years gained than the next less intensive efficient strategy. Strategies on this line would be considered efficient because they achieve the greatest gain in benefit (life years gained) per harm or use of mammograms. Strategies that use more mammograms but still have small benefits (i.e., a shallower slope than the next best strategy) are considered to be less efficient (i.e., weakly dominated). When and if the slope in the efficiency frontier plot levels off, it means that the additional life-years gained per increase in mammography are small relative to the previous strategies and could indicate a point at which additional screening might be considered as having a low return (or additional benefit). There is no definitive inflection point across the models for the strategies or metrics evaluated. Black strategies are efficient; gray strategies close to the efficiency frontier are less efficient; and open gray strategies are inefficient (inferior, or dominated). Reference 33 provides efficiency frontiers for other harm and benefit metrics. D = Dana-Farber Cancer Institute; E = Erasmus Medical Center; GE = Georgetown University Medical Center and Albert Einstein College of Medicine; M = MD Anderson Cancer Center; QALY = quality-adjusted life-year; S = Stanford University; W = University of Wisconsin and Harvard Medical School.