Technical Report

Collaborative Modeling of U.S. Breast Cancer Screening Strategies

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850

Prepared by:
Writing Committee of the Breast Cancer Working Group
Cancer Intervention and Surveillance Modeling Network (CISNET) and the Breast Cancer Surveillance Consortium (BCSC)

Writing Committee Members:
Jeanne Mandelblatt
Kathy Cronin
Harry de Koning
Diana L. Miglioretti
Clyde Schechter
Natasha Stout

AHRQ Publication No. 14-05201-EF-4
December 2015

The modeling analysis included in this report was done by six independent teams from the Dana-Farber Cancer Institute (PI: Lee); Erasmus University (PI: de Koning); Georgetown University Medical Center, Lombardi Comprehensive Cancer Center/Albert Einstein College of Medicine (PI: Mandelblatt/Schechter); University of Wisconsin/Harvard Medical School, Harvard Pilgrim Health Care (PI: Trentham-Dietz/Stout); M.D. Anderson Comprehensive Cancer Center (PI: Berry); and Stanford University (PI: Plevritis).

This work was supported by the National Institutes of Health under National Cancer Institute Grant U01 CA152958 and the National Cancer Institute-funded Breast Cancer Surveillance Consortium (BCSC) Grant P01 CA154292, contract HSN261201100031C, and Grant U54CA163303. The collection of BCSC cancer and vital status data used in this study was supported in part by several state public health departments and cancer registries throughout the United States. For a full description of these sources, go to: http://breastscreening.cancer.gov/work/acknowledgement.html.

Model results and the contents of this report are the sole responsibility of the investigators.

Authors and Affiliations

Jeanne S. Mandelblatt, Aimee M. Near, Amanda Hoeffken, and Yaojen Chang: Department of Oncology, Georgetown University Medical Center and Cancer Prevention and Control Program, Georgetown-Lombardi Comprehensive Cancer Center, Washington, DC

Natasha K. Stout: Department of Population Medicine, Harvard Medical School, and the Harvard Pilgrim Health Care Institute, Boston, MA

Clyde B. Schechter: Departments of Family and Social Medicine and Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY

Harry J. de Koning, Nicolien T. van Ravesteyn, Jeroen J. van den Broek, and Eveline A. Heijnsdijk: Department of Public Health, Erasmus Medical Center, Rotterdam, the Netherlands

Diana L. Miglioretti: Department of Public Health Sciences, University of California, Davis School of Medicine, Davis, CA, and Group Health Research Institute, Group Health Cooperative, Seattle, WA

Martin Krapcho: Information Management Services, Calverton, MD

Amy Trentham-Dietz, Oguzhan Alagoz, and Ronald Gangnon: Carbone Cancer Center, University of Wisconsin, Madison, WI

Diego Munoz: Departments of Biomedical Informatics and Radiology, School of Medicine, Stanford University, Stanford, CA

Sandra J. Lee and Hui Huang: Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, and Harvard Medical School, Boston, MA

Donald A. Berry, Gary Chisolm, and Xuelin Huang: Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX

Oguzhan Alagoz and Mehmet Ali Ergun: Department of Industrial and Systems Engineering, University of Wisconsin, Madison, WI

Karla Kerlikowske: Departments of Medicine and Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA

Anna N.A. Tosteson: Norris Cotton Cancer Center and The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine, Dartmouth University, Hanover, NH

Ronald Gangnon: Department of Biostatistics and Medical Informatics and Population Health Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI

Brian L. Sprague: Department of Surgery, College of Medicine, University of Vermont, Burlington, VT

Sylvia Plevritis: Department of Radiology, School of Medicine, Stanford University, Stanford, CA

Kathleen A. Cronin and Eric Feuer: Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD

Acknowledgments

The authors thank Jennifer Croswell from the Agency for Healthcare Research and Quality; members of the U.S. Preventive Services Task Force; the Oregon Evidence-based Practice Center; Jessica Garshell for data processing; and Adrienne Ryans for report preparation. The authors also thank Paul Pinsky, Stuart Baker, William Lawrence, Tom Trikalinos, and John Wong for helpful suggestions on earlier versions of this report.

Executive Summary

This report summarizes the methods and results of simulation modeling of alternative digital mammography breast cancer screening strategies for the U.S. female population. Six established simulation models from the Cancer Intervention and Surveillance Modeling Network and investigators from the Breast Cancer Surveillance Consortium were commissioned by the U.S. Preventive Services Task Force to evaluate the benefits and harms of strategies that varied by age of screening initiation and cessation and screening intervals for average-risk women. In secondary analyses, we assessed how disutility related to the screening process and its consequences affected the balance of benefits and harms of the different screening strategies. Additionally, we conducted analyses to examine how the balance changed if the screening approach considered risk for breast cancer, breast density, or comorbidity. Finally, we conducted sensitivity analyses and analyses validating the models.

The models portray four molecular subtypes of breast cancer based on hormone receptor and human epidermal growth factor-2 receptor status. They used a lifetime perspective and common data inputs on incidence, risk and breast density prevalence, digital mammography performance, treatment effects, and other-cause mortality among a cohort of women born in the United States in 1970. The specific strategies assessed included screening beginning from ages 40, 45, or 50 years to age 74 years at annual or biennial intervals, or annually from ages 40 to 49 or 45 to 49 years and biennially thereafter. All strategies are compared to the counterfactual situation of no screening; strategies are all compared incrementally to each other. To evaluate program efficacy, all analyses assumed 100 percent screening adherence and used subtype-specific, guideline-recommended systemic treatment. Outcomes were considered over the entire lifetime of the cohort. There were several benefit outcome metrics considered, including the percent reduction in breast cancer mortality (vs. no screening); breast cancer deaths averted; life-years gained; and quality-adjusted life-years gained. Harms included number of mammograms; false-positives; and benign biopsies. Another metric was overdiagnosis, defined as cases that would not have been clinically detected in a woman's lifetime in the absence of screening because of lack of progressive potential or death from competing mortality. This was operationally calculated in the models by subtracting the total number of cases in a screening scenario from the total number of cases diagnosed in the absence of screening.

In validation analyses, the models reproduced results of trends in observed U.S. incidence and mortality as well as 13-year followup results from the U.K. trial on screening women annually in their 40s.

In an unscreened population, the models predict a median 12.9 percent cumulative probability of developing breast cancer from ages 40 to 100 years (range across models, 12.0% to 14.0%). Without screening, the median probability of dying of breast cancer is 2.5 percent (range, 1.5% to 3.2%). Thus, if a particular screening strategy leads to a 30 percent reduction in breast cancer mortality, then the probability of breast cancer mortality would be reduced from 2.5 to 1.8 percent, or 7.5 breast cancer deaths averted per 1,000 women screened.

The six models produced consistent rankings of the strategies evaluated. Screening biennially from ages 50 to 74 years achieves a median 25.8 percent (range, 24.1% to 31.8%) breast cancer

mortality reduction versus no screening, and averts 7.1 (range, 3.8 to 8.7) breast cancer deaths. Biennial strategies maintain an average 81.2 percent (range across strategies and models, 68.3% to 98.9%) of annual screening benefits with almost half the false-positives and fewer overdiagnosed cases. Compared to biennial screening from ages 50 to 74 years, annual screening from ages 40 to 74 years reduces mortality an additional 12.0 percent (range, 5.7% to 17.2%) and averts 3 more breast cancer deaths, but yields 1,988 more false-positives and 11 more overdiagnoses per 1,000 women screened. Of note, compared to alternative strategies, annual screening from ages 50 to 74 years is consistently dominated (that is, uses more resources but has less benefit) across all outcome metrics, and would be considered inefficient. This is because there is only a small added incremental benefit of annual versus biennial screening in the 50- to 74-year-old age group, while annual screening requires twice as many mammograms and generates nearly double the number of false-positives.

If disutility of screening, having a false-positive, living with cancer, and having decrements in general health with aging are considered, the ranking of strategies changes, although the benefits are reduced (i.e., quality-adjusted life-years are lower than life-years). Screening continues to have benefits, albeit smaller, in all age groups, including those ages 40 to 49 years. Sensitivity analyses examining a range of disutility values did not change conclusions of the analyses.

We specifically examined risk levels applied starting from age 40 years until death that were 1.3-, 2-, and 4-fold higher than average, corresponding to women in their 40s who have heterogeneously or extremely dense breasts, a family history of a first-degree relative with breast cancer (excluding risk for *BRCA1* and *BRCA2* gene mutations), or a combination of family history and other risk factors, respectively. The ranking of strategies does not change when screening is based on risk levels; annual screening from ages 50 to 74 years remains dominated by other approaches. However, the balance of benefits and harms over a range of risk groups differs, with women who have higher risk obtaining greater gains from screening and experiencing lower rates of false-positives than women in the lowest-risk groups. Screening higher-risk women also yields a lower proportion of overdiagnosed cases per death averted than screening women of average population risk. For women in their 40s with a 2- to 4-fold increase in breast cancer risk compared to the average population in their age group, annual screening starting at age 40 or 45 years would have a similar or more favorable harm to benefit ratio (based on false-positives per death averted) than biennial screening in average-risk women from ages 50 to 74 years. For women with even a 1.3-fold increase in risk, biennial screening starting at age 40 years would have a similar ratio of harms to benefits as biennial screening of average-risk groups from ages 50 to 74 years. Results are generally similar for ratios of harms to benefits based on other outcome metrics.

Considering breast density alone or in combination with other risk factors does not affect the ranking of strategies, and annual screening from ages 50 to 74 years continues to be dominated for all breast density groups. For women with no comorbidity who have an average of a 17-year remaining life expectancy, screening would be efficient through ages 78 to 80 years and would have a minimal increase in overdiagnosis compared to stopping at age 74 years. However, for women with moderate to severe comorbidity, screening cessation could occur at about age 68 years.

While the models produced very consistent results in the ranking of screening approaches, there are acknowledged limitations to the modeling analysis. First, there is expected variability across models in estimates of benefits and harms based on differences in model structure and assumptions. The models have the greatest variability in results for overdiagnosis, since this is an unobservable phenomenon for which there are currently no primary biologically-based data. Thus, overdiagnosis must be inferred indirectly. Many methods for this have been proposed, and there is no gold standard approach. Modeling makes a useful contribution to estimating overdiagnosis since it explicitly considers lead-time and competing mortality and takes a lifetime perspective. Overall, using multiple models produces a range of results for overdiagnosis (and other screening outcomes) that can be useful to decisionmakers.

Second, the modeling did not consider other imaging technologies, polygenic risk, or risk of breast cancer related to screening. Also, we assumed 100 percent adherence to screening, prompt evaluation of abnormal results, and full use of optimal treatment to evaluate program efficacy. Benefits will always fall short of the projected results since access and use is not universal. Finally, these analyses were designed to provide modeling data for use in public health decisionmaking for populations of women; the results are not intended to guide individual screening decisions.

In summary, from the vantage point of public health programs developed for the overall U.S. female population, the six models produce a consistent ranking of the various breast cancer screening strategies and conclude that biennial screening strategies are most frequently the most efficient. All six modeling groups also project some benefits associated with screening women starting at age 40 years, and while screening initiation at age 40 years has the greatest benefits, it also has the greatest harms. Thus decisions depend on tolerance for additional false-positives, biopsies, and overdiagnosed cases. The ranking of strategies is not affected by risk level or breast density; however, annual screening of women ages 40 to 74 years with a 2- to 4-fold increased risk or biennial screening of those with a 1.3-fold increased risk has a comparable ratio of benefits to harms as biennial screening from age 50 to 74 years in the average-risk population. Among women with severe or moderate levels of comorbidity, harms of screening seem to outweigh benefits prior to age 74 years, but for those with no or mild levels of comorbidity, screening benefits continue to age 78 to 80 years, with minimal increases in overdiagnosis. Choices about optimal ages of initiation and cessation will ultimately depend on program goals, weight attached to the balance of harms and benefits, and considerations of efficiency.

Table of Contents

Figures

Tables

Chapter 1. Introduction

We used six established simulation models to synthesize data and evaluate multiple digital mammography screening strategies in the U.S. population (1-3). Modeling has the advantage of providing a quantitative summary of the net balance of harms and benefits and considering preferences (utilities) while holding selected conditions (e.g., screening intervals or test sensitivity) constant, facilitating strategy comparisons. Because all models make assumptions about unobservable events, collaboration of several models provides a range of plausible effects and can illustrate the effects of differences in model assumptions (2-4).

In this report, we summarize model methods, data sources, and results and discuss the strengths and limits of our approach to evaluating screening with digital mammography from ages 40, 45, or 50 years to age 74 years at different intervals among average-risk women. In secondary analyses, we also examined how breast density and risk or comorbidity level affects results, and whether utilities for health states related to screening and its downstream consequences affect conclusions.

Chapter 2. Methods

The breast cancer models were developed within the Breast Working Group (BWG) of the Cancer Intervention and Surveillance Modeling Network (CISNET) of the National Cancer Institute (5-11) and were exempt from institutional review board approval. As the oldest, longest-funded CISNET group, the BWG has demonstrated the value of collaborative modeling. The six models were independently developed to examine the impact of breast cancer control interventions on population trends in incidence and mortality, but they share common features, including: 1) following multiple birth cohorts over time, 2) incorporating known data on breast cancer biology, 3) using common data about screening behavior and treatment use based on known accuracy or effectiveness, and 4) projecting future benefits. The six BWG groups consist of scientists from complementary disciplines, including actuarial science, biostatistics, economics, epidemiology, industrial and systems engineering, health services and health policy research, medicine, and oncology. The groups are joined by key national partners to ensure that the modeling research reflects state-of-the-art knowledge and available data and can be readily disseminated. Over the past 14 years, the BWG has been highly productive, collectively publishing 162 manuscripts, including those that have engaged the research and policy communities in collaborative modeling activities that have had a direct public health impact (2, 3, 12).

The models in the CISNET BWG include Model D (Dana-Farber Cancer Institute, Boston, MA), Model E (Erasmus Medical Center, Rotterdam, the Netherlands), Model G-E (Georgetown University Medical Center, Washington, DC, and Albert Einstein College of Medicine, Bronx, NY), Model M (M.D. Anderson Cancer Center, Houston, TX), Model S (Stanford University, Stanford, CA), and Model W (University of Wisconsin, Madison, WI, and Harvard Pilgrim Health Care, Boston, MA).

Each model portrays four distinct molecular subtypes, each with its own trajectories and responses to therapy, based on estrogen receptor (ER) and human epidermal growth factor-2 receptor (HER2) status (4). The models have been recently updated to reflect current population trends in incidence (13, 14) and competing nonbreast cancer mortality. Screening performance reflects modern digital technology and the most current therapeutic trial results. All models except one (Model S) includes ductal carcinoma in situ (DCIS); Model S only portrays invasive cancer. The general modeling approach is summarized below, followed by specific details about each model.

The models begin with estimates of overall breast cancer incidence and ER/HER2-specific survival trends *without* screening or adjuvant treatment and then overlay data on screening use and reductions in mortality associated with adjuvant treatment for each molecular subtype to generate observed U.S. population incidence and mortality trends; the models assume that all women diagnosed with breast cancer receive local treatment (2-4, 15-18). Women are assumed to have average risk; risk levels, including risk associated with breast density, can modify incidence. Each breast cancer is depicted as having a distribution of preclinical screening-detectable periods (sojourn time) and a clinical detection point. Age, screening round and interval, and breast density affect mammography performance. On the basis of digital

mammography sensitivity (or thresholds of detection), screening identifies disease in the preclinical screening-detection period and results in the identification of earlier stage or smaller tumors than might occur via clinical detection, resulting in local and systemic treatment with a corresponding reduction in breast cancer mortality. At the time of diagnosis, ER/HER2 status is assigned based on stage and age. Molecular subtype–specific treatment reduces the hazards of breast cancer death (Models D, G-E, M, and S) or results in a cure for some fraction of cases (Models E and W). Women can die of breast cancer or other causes.

Dana-Farber (Model D)

Model D is a stochastic model that depicts the early detection process of screening and predicts breast cancer mortality as a function of the disease natural history, detection process, and treatment (9, 19). Model D takes an analytical approach to estimate the impact of mammography screening and treatment on incidence and mortality of breast cancer. The factors that influence mortality in Model D are: examination schedule and sensitivity, transition into the preclinical and clinical states, distribution of the preclinical sojourn time, age distribution of the population, length of followup, incidence of disease by age and other risk factors, and unique factors associated with the natural history of the disease.

Model D characterizes the natural history of invasive breast cancer by four health states: S0 is a disease-free state (disease-free or disease cannot be detected by any screening modality), Sp is a preclinical state (disease can be diagnosed by a screening test), Sc is a clinical state (symptomatic disease), and Sd is a disease-specific death state. There are two main model assumptions: 1) invasive breast cancer is progressive and described by the transitions from S0 to Sp to Sc, and some eventually progresses to Sd; and 2) the mortality benefit from screening is from a stage shift in diagnosis. The main goal of screening is to diagnose individuals in Sp, where subjects have an early-state disease with no symptoms. The stage shift implies the subjects are diagnosed earlier (in Sp), before symptoms surface (in Sc). Model D mathematically derives a distribution of the lead-time in the presence of screening and adjusts the lead-time bias in mortality modeling. The model assumptions have been validated by projecting the results from randomized screening trials and comparing model outputs to published trial results (20).

Since 2009, a second potential path of DCIS was incorporated into Model D, as shown in **Figure 1**. The revised Model D envisions that normal tissue can progress to either early-stage DCIS or preclinical invasive breast cancer. Invasive breast cancer follows the health states described above. For early-stage DCIS (preclinical DCIS), Model D assumes it can potentially take one of these three paths: 1) stay in the early stage of DCIS and/or eventually regress to normal; 2) progress to invasive breast cancer; or 3) progress to a later stage of DCIS (clinical DCIS), where clinical symptoms appear. It is assumed one does not die of DCIS. Furthermore, it is assumed that mammography screening finds individuals with preclinical DCIS. For preclinical DCIS that will eventually progress to clinical DCIS (i.e., path 3), the mean sojourn time was estimated to be 2 to 3 years, with exponential distribution using DCIS incidence data from the Surveillance, Epidemiology and End Results (SEER) program from 1973 to 1979 (i.e., the prescreening era). The transition probability from early- to late-stage DCIS, W1b(t), was estimated using age-period-cohort (APC)-based DCIS incidence data in the absence of screening for the 1970 birth

cohort. A net transition probability of [W1a(t)-Wr(t)- W1c(t)] to the reservoir of early-stage DCIS (entering–leaving the reservoir) was estimated using the method described by Lee and Zelen (21).

The main assumption in Model D is that mortality benefits from screening are due to a stage shift by finding disease earlier, when prognosis is more favorable. For example, approximately 50 percent of women diagnosed in the clinical state (usual care) tend to be in a node-negative stage compared to 70 to 80 percent of women diagnosed by an early detection program (e.g., mammography). Therefore, the survival distribution (conditional on disease stage at early diagnosis) will be more favorable for women whose cancer is screen detected versus clinically detected. Model D utilizes the stage distribution data by the mode of detection provided by BCSC. Model D incorporates the probability distribution of the lead-time in adjusting mortality for screen-diagnosed cases.

Model D derived the lead-time distributions for DCIS and invasive breast cancer in the presence of screening. Using these distributions, the probability of overdiagnosis conditional on being screen-detected (i.e., lead-time is longer than residual survival time) was estimated for DCIS and invasive breast cancer. These probabilities were applied to the screen-detected cases to quantify overdiagnosis. This method of estimating overdiagnosis was compared to the difference between total number of diagnosed cases in the presence and absence of screening. There was good agreement.

The survival benefits of various adjuvant therapies were assessed based on a meta-analysis of clinical trial results (22). The reported estimates on proportional reduction in the annual odds of death for treatment by age groups and ER/HER2 status were applied to the baseline survival. The baseline survival in the absence of treatment was assessed using SEER breast cancer–specific survival data from cases diagnosed in 1975 to 1979. For the 1970 birth cohort, the best available treatment benefit by ER/HER2 status was applied.

Erasmus (Model E)

Model E is called the MISCAN-Fadia (MIcrosimulation of SCreening Analysis–Fatal diameter) model. It is a microsimulation model generating independent life histories. MISCAN-Fadia is unique in that it explicitly models invasive tumor growth in combination with the concept of a fatal diameter (10). The model also includes DCIS. Model E simulates a large population of women using the demographic characteristics of the U.S. female population. The simulated population consists of individual life histories, in which some women develop breast cancer and some die of the disease. A certain percentage of the modeled population develops preclinical disease. This percentage varies between birth cohorts and is based on cumulative incidence (23). The cohorts have the same age distribution of onset of breast cancer. The age distribution of onset is based on 1975 age-specific incidence rates, with a shift to younger ages, because onset of tumor growth was earlier than clinical diagnosis in the 1975 prescreening era. The simulated woman dies of breast cancer or of other causes, whichever comes first. The following sections include a summary of the structure and assumptions for each component of the model.

Among those who develop disease, the natural history of breast cancer is modeled as a continuously growing tumor. Each tumor has a size (the fatal diameter, which differs between tumors) at which diagnosis and treatment will no longer result in cure given available treatment options. If the tumor is diagnosed (either on the basis of clinical presentation with symptoms or by screening) and treated before it reaches the fatal diameter, the woman will be cured (and will die of nonbreast cancer causes). Variation between tumors is modeled by probability distributions of tumor growth, threshold diameter of screen detection, clinical diagnosis diameter, and fatal disease diameter. The tumor growth rate, survival duration since fatal diameter, threshold diameter for screen detection, clinical detection diameter, clinical diagnosis because of distant metastases, and correlations are estimated using data from the Two County Study (24, 25). The fatal diameter was calibrated to U.S. data based on 1975 stage distribution and survival (26). Survival is modeled from the time of fatal diameter. For observed survival, the time between clinical diagnosis and the time the tumor reached the fatal diameter is subtracted.

MISCAN-Fadia includes a submodel for DCIS. DCIS can either regress, become invasive, or be clinically diagnosed at exponential rates. These rates are estimated using SEER American Joint Committee on Cancer (AJCC) stage- and age-specific incidence rates for DCIS and invasive cancer from 1975 to 1999. For example, the rate at which DCIS becomes clinically diagnosed is based on the small percentage of DCIS that was diagnosed prior to use of screening in 1975 to 1979. This natural history approach readily lends itself to defining separate distributions for each of these parameters based on risk groups and molecular tumor subtype.

When a screening program is applied, the preclinical tumor may be detected by screening. Each simulated tumor has a diameter at which it will be clinically diagnosed and a threshold diameter of screen detection. For the latter, screening test sensitivity is 0 percent below and 100 percent above this diameter. The threshold diameter depends on the calendar year and age of the woman (decreasing with calendar year and older age). Screening benefits result from detection of more tumors at a nonfatal size. The dissemination of mammography is modeled based on the actual dissemination in the U.S. population (27). In addition, specified screening programs (with fixed screening intervals and starting and stopping ages) can be incorporated in the model. This structure provides flexibility to define different thresholds of detection for any screening test based on molecular subtypes of cancer.

Model E simulates life histories for individual women. The model uses the so-called parallel universe approach and first simulates the individual life histories for women in the absence of screening and then assesses how these histories would change as a consequence of a screening. To estimate the amount of overdiagnosis, the number of breast cancers detected in these two situations are compared. Overdiagnosis is defined as “the detection of tumors that would not have been detected in a woman’s lifetime in the absence of screening.” Overdiagnosis can occur because of lack of progressive potential (e.g., of DCIS) or because a woman dies from another cause before the breast cancer would have been clinically detected. The amount of overdiagnosis is calculated by subtracting the number of breast cancer diagnoses in the absence of screening from the number of breast cancer diagnoses in the presence of screening.

The benefit of adjuvant treatment is modeled as a shift in the fatal diameter for treated women. For each adjuvant treatment, a cure proportion is estimated (depending on age) using treatment

effectiveness data (based on meta-analyses of the Early Breast Cancer Trialists' Collaborative Group) (22). These cure proportions are then translated into corresponding fatal diameters (i.e., a more effective treatment can cure a larger tumor).

Georgetown-Einstein (Model G-E)

Model G-E is a microsimulation of breast cancer in the U.S. population, implemented in C++ programming language, that is specifically oriented toward estimating the impact of screening and adjuvant treatment innovations that have taken place since 1975 (6). The approach is phenomenological; there is no attempt to model any specific biology of breast cancer. The impact of screening and treatment are managed by creating "parallel universes," whereby the same life history is subjected to different real or counterfactual screening or treatment strategies and the varying results are directly compared. The model's inputs have been calibrated to produce a reasonable approximation to SEER incidence and mortality over the period of 1975 to 2010. The mortality risk conferred by any given breast cancer depends upon the biomarkers, the patient's age at diagnosis, the stage at diagnosis, and the treatment provided.

Breast cancer is assumed to exist in two forms: progressive and nonprogressive. Nonprogressive lesions have a transient existence and are never identified clinically, but may be detected through screening and present as DCIS when they are. Nonprogressive breast cancer has no mortality associated with it. Progressive lesions may present clinically or through screen detection, in any of the AJCC stages, and all of these lesions carry a risk of breast cancer mortality. All breast cancers, progressive or nonprogressive, may be classified by the presence or absence of two biomarkers: ER and HER2. The incidence of breast cancer depends on a woman's birth cohort and varies with age. The age-specific incidence rates, in turn, depend on the woman's breast density. The mortality risk conferred by any given breast cancer depends upon these biomarkers, the patient's age at diagnosis, the stage at diagnosis, and the treatment provided.

In the simulation, construction of a life history begins by selecting a birth cohort for each woman, sampled from the distribution of population birth years from U.S. Census data, or, in some applications, a single birth cohort is simulated. A date of death from nonbreast cancer causes is sampled from a birth cohort–specific life table.

Incidence of breast cancer in the (counterfactual) absence of screening is based on a modification of an APC model (23), extended beyond its covered ages and cohorts by applying year-on-year incidence ratios (13, 14) and then further calibrated to improve match to SEER incidence from 1975 to 2010. A time-to-event distribution for onset of clinical breast cancer is sampled to determine when, if ever, the woman will develop clinically apparent breast cancer. If clinically apparent breast cancer will develop, it is assigned a stage by sampling the age-specific stage distribution for clinically detected cancer, and is then given a biomarker classification by sampling the biomarker distribution conditional on age and stage. Survival from the time of clinical diagnosis in the absence of treatment is then sampled from a time-to-event distribution conditional on age, stage, and biomarkers using the survival functions describing prognosis of breast cancer in 1975, and the corresponding date of death from breast cancer (which may be before or after the date of nonbreast cancer death) is calculated. Finally, a sojourn time for the

lesion is sampled. The sojourn time distributions are conditional on age at clinical presentation and biomarkers. All of the conditional distributions are assumed to be gamma distributions with a common shape parameter. (The value of the shape parameter is an input, as are age-biomarker–specific means.) A date preceding the date of clinical onset of breast cancer by the duration of the sojourn time is identified as the onset of the sojourn period for this lesion. If no clinically apparent breast cancer is to develop, time-to-event distributions for onset and regression of nonprogressive lesions are sampled to determine when, if ever, such a lesion develops. Parameters of these distributions are among those calibrated to produce a match to SEER incidence after the dissemination of screening in the United States. The nonprogressive lesion is assigned an ER/HER2 classification by sampling from the biomarker distribution of all DCIS lesions for a woman of her age, and its stage is set to DCIS.

The above steps create a basic life history describing breast cancer in the absence of screening or adjuvant treatment, characterizing each simulated woman by a birth date, date of death from nonbreast cancer causes, and, in women with breast cancer, dates of sojourn onset, clinical presentation, and death from breast cancer.

Each woman is assigned to a mammography screening schedule (or, in simulations including counterfactual screening strategies, several screening schedules). The dissemination screening schedule is randomly sampled to produce birth cohort–specific screening schedules that are thought to resemble actual screening behavior among women in the United States. Counterfactual, strictly periodic screening schedules, such as "every 2 years from ages 50 to 74 years," can be simulated as well, as can scenarios in which screening intervals vary with age or breast density. If a screening mammogram is carried out during the sojourn interval of a breast cancer, there is a probability that it will be detected. This probability, known as the sensitivity, depends on the woman's age at the time of the screening; whether it is an initial or subsequent screening; in some analyses, the woman's breast density; and whether the mammogram uses film or digital technology. Note that this sensitivity is an abstract, unobservable parameter of the model that is calibrated to reproduce 1-year screen-detection hit rates from BCSC. The actual outcome of the simulated screening is determined by sampling a uniform random number and comparing that to the sensitivity. If screen detection occurs, a new stage, possibly earlier than the clinical stage, is assigned to the lesion. To do this, the model draws on distributions of stage dwell. The distributions are assumed to be exponential, and the means are unconditional program inputs. Based on the lead-time obtained by screening and the dwell time distributions, a screened stage is sampled from a Bayesian posterior distribution. Survival in the absence of adjuvant treatment is then recalculated based on the new age and stage at diagnosis (the biomarkers are assumed to be the same as if the lesion were diagnosed clinically), and the date of death from breast cancer is revised accordingly.

In light of the above, the effect of screening on breast cancer mortality is based entirely on stage shift and age shift. Once a lesion has been screen detected, screening terminates. If a lesion goes undetected at every screening examination, it will still present clinically at its clinical presentation date (unless it is nonprogressive, in which case it will eventually regress). Screening examinations conducted before the sojourn period or in a woman with no breast cancer in her life history have a probability of leading to a false-positive result. This probability is 1 minus the specificity of the test. Test specificity is conditional on age; whether it is an initial or subsequent

screening; screening technology; and, in certain analyses, breast density. False-positive screening tests do not interrupt the screening schedule.

As described above, nonprogressive lesions have a transient existence and are never identified clinically, but may be detected through screening and present as DCIS when they are. DCIS can either regress, become invasive, or be clinically diagnosed at a specified nonexponential distribution. Model G-E assumes that all invasive disease can progress to lethality; therefore, invasive overdiagnosis arises only as a result of other-cause mortality intervening between screen detection and the date when the lesion would present clinically.

Upon clinical diagnosis or screen detection, a woman with a breast cancer diagnosis is assigned a treatment. In the basic model, this is done by sampling from a distribution of treatments specific to age, stage, year of diagnosis, and biomarker. These distributions are program inputs thought to represent the dissemination of adjuvant therapies in the U.S. population. Counterfactual treatment distributions (e.g., every woman receives the most effective treatment available at the time for her age and biomarker combination) are also available. Each combination of treatment and lesion characteristics (age at diagnosis, stage, and biomarkers) is associated with a hazard ratio less than or equal to 1, which specifies the treatment effectiveness. Although the model is programmed to also apply cure fractions in association with treatment, all implementations of the model so far have assumed that all cure fractions are zero and have relied exclusively on hazard reduction. The survival curve for the lesion, with the treatment-associated hazard ratio applied, is sampled to determine a new survival duration, and the date of death from breast cancer is modified accordingly.

Model G-E has been through one major and several minor revisions since its first appearance in 2001. Most of the changes have been enhancements to the code and its numerical and sampling algorithms to make it run more efficiently, including a change from modular coding in C in the original, to object-oriented programming in C++ in 2006 to 2007, to an upgrade from ANSI standard C++ to ISO standard C++ in 2012 to 2013. These changes, along with improvements in hardware and operating systems, have greatly enhanced the scope of simulations. Whereas simulation of a single scenario for N=50,000,000 required more than 24 hours of computing time in the original version, we can now simulate N=200,000,000 for 15 parallel scenarios in approximately 2 hours.

A number of substantive changes to the program, reflecting the modelers' emerging perspectives on breast cancer, have also occurred. These are briefly summarized here.

1. The original program simulated only a single scenario (combination of screening and treatment strategies) at a time and maintained all events in an event queue. There was only minimal parallelism between the life histories generated in different scenarios. When the model was reprogrammed in 2006, the event queue was eliminated, and all relevant dates were maintained in a life-history object. Since then, any number of scenarios can be processed in parallel, and the same underlying life histories are used for all scenarios. In addition, depending on the nature of the scenario, it is sometimes possible to impose parallelism among different screening programs or treatment plans.

2. Natural histories were not distinguished by biomarker categories in the original model (although responses to treatment were). These former were added in 2011.
3. Sojourn time distributions were all assumed to be exponential in the original model; this constraint to the gamma family was relaxed in 2012 when it became apparent that the model was overgenerating lesions with extremely long or short sojourn times.
4. Density-specific natural histories and screening operating characteristics were added in 2012.
5. Nonprogressive disease was added in 2006.
6. The original simulation output provided only counts of mammograms, incident cases in each stage, breast cancer deaths, and surviving population for each age and calendar year combination. Numerous additional outputs have been added, including counts of nonprogressive cases diagnosed, overdiagnosed cases, and life-years (overall and in specific stages of breast cancer treatment).

In addition to the structural model changes noted above, parameter estimates have changed over time, driven by the awareness of new knowledge and the need to calibrate the incidence outputs to more recent SEER data. Specific Model G-E changes include:

1. While all CISNET models have adopted a new APC model as the basis for estimating breast cancer incidence in the absence of screening, several variants of that model are in use. We begin with the Holford APC model (23) and extend it to cover older ages and more recent birth cohorts by applying year-on-year incidence ratios (14). We then add an additional period effect that rises from 1987 through 1997 and then declines again through 2007.
2. Nonprogressive disease incidence is calculated as a fixed proportion of all DCIS. (Previously this was calculated as a fixed proportion of all breast cancer.)
3. Mammography dissemination is slightly modified from the Cronin-Krapcho model (27) to reduce the amount of mammography use by women born before 1948 during the first half of the 1980s.
4. The sojourn period of nonprogressive lesions is 6 years up until 1995. Reflecting increased efforts by radiologists to detect ever smaller lesions with plain film mammography, the sojourn times increase linearly between 1995 and 2005, topping out at 12 years and remaining level thereafter. The increased sojourn time of nonprogressive lesions is achieved by allowing them to become detectable at an earlier point in their lifespan.
5. Treatment hazard ratios are based on earlier estimates from the Early Breast Cancer Trialists' Collaborative Group meta-analysis (22) rather than the more recent treatment dissemination parameters that some other CISNET models use, since the former reflects greater treatment effects over time that are more consistent with changes in clinical practice.

M.D. Anderson (Model M)

Model M is a Bayesian simulation model (7). It simulates a population of 1 million women that has the age distribution of the United States in 1975. For each virtual woman, the model simulates a natural course of her life separate from breast cancer. Each year, each woman is diagnosed with breast cancer or not, depending on the incidence of the disease for women her age in that year, whether she had a screening mammogram that year, and her history of mammography. In the absence of screening, Model M assumes that incidence is the same as it

was in the prescreening era based on rates reported to SEER from 1975 to 1979, with an essentially flat trend over time; no cohort trend is included. The model keeps track of which women are diagnosed with breast cancer in each year and which women die of breast cancer and from other causes. The model considers births and deaths.

Model M is not a natural history model. S(t) denotes the probability of surviving breast cancer to time t after diagnosis for cancer detected in the absence of both screening and adjuvant therapy. S(t) depends on age, stage, and ER/HER2 status. Model M uses the standard CISNET estimate for S(t) (4) but allows for uncertainty in this estimate by incorporating an unknown parameter h. This parameter affects the hazard function of S(t) multiplicatively so that the resulting new survival function is S(t) raised to the power of h. Consistent with the Bayesian approach, Model M assesses prior probability distributions for all unknown parameters. The parameters of interest are those that affect the diagnosis of breast cancer (screening) and its course once the disease is diagnosed (treatment). The joint prior distribution of these parameters is updated using Bayes' rule and based on SEER breast cancer mortality over time. The calculation process is to first generate a vector of values from the joint prior distribution of the parameters. Together with the known inputs, these parameters are sufficiently specific to enable generation of the breast cancer history of each woman in the entire virtual population, including whether and when breast cancer is the cause of death. This vector of parameter values is accepted as an observation from the joint posterior distribution if the simulated breast cancer mortality over time is sufficiently close to SEER breast cancer mortality. This process is repeated many times to accumulate enough acceptances to allow for adequately estimating the posterior distribution. This posterior distribution is used to make inferences about the unknown parameters and to produce additional simulations assuming hypothetical screening histories, treatments, and treatment effects.

Screening helps detect cancer in an earlier stage (stage shift). Screen-detected cases are also less likely to result in breast cancer death than similar clinically-detected cases. This is called "beyond stage shift." Screening parameters also include a cure fraction for AJCC stage I and II cancers.

Overdiagnosis is the difference between the number of cases of breast cancer when there is the indicated amount of screening versus no screening. Treatment parameters include those for the effects of combinations of adjuvant hormone therapy; adjuvant trastuzumab; adjuvant therapy with cyclophosphamide, methotrexate, and fluorouracil; adjuvant anthracycline-based therapy; and adjuvant taxanes. Treatment affects mortality by a hazard reduction and cure fraction by mode of diagnosis.

Stanford (Model S)

Model S is a multibirth cohort model that simulates the impact of breast cancer screening and adjuvant treatment on U.S. breast cancer incidence and mortality trends (8). Model S generates a series of breast cancer–specific events in the life history of individual breast cancer patients, then aggregates these patient-specific events to produce population-level trends. For each individual patient's life history, Model S generates her age at detection, mode of detection (screening vs. symptoms), tumor size, survival time, and cause of death (breast cancer vs. other causes).

Moreover, Model S describes the preclinical progression of each patient's disease prior to symptoms in order to determine the likelihood of detection by screening and the smaller tumor size attributed to screen detection. Each simulated event in an individual patient's history is a random variable, drawn from a probability distribution function.

The natural history of invasive breast cancer is modeled as progressive disease, where an individual patient's tumor grows exponentially in size and has an increasing probability of advancing to regional and distant stages (or the comparable AJCC stages) at larger sizes (8, 28). Each simulated patient is randomly assigned a tumor volume-doubling time, drawn from a distribution that is conditional on her age at clinical detection in the absence of screening. Younger women have faster growing tumors than older women. The tumor's history is reconstructed "backward" from the time of symptomatic detection to the moment it had been 2 mm in diameter, as opposed to being reconstructed "forward" from initiation of the first malignant cell. Progression from preinvasive disease, such as DCIS, to invasive disease is not modeled, but could be included in future work. As such, Model S is the only BWG model that does not include DCIS. Model S incorporates the use of underlying breast cancer–specific survival by ER/HER2 subtype, as well as its predicted distribution at clinical detection in the absence of screening.

Breast cancer incidence in the absence of screening is estimated using the APC model commonly used among modeling groups (14, 23). However, contrary to other groups, Model S does not rely simply on this input. Instead, Model S uses background breast cancer incidence derived from a novel approach developed under the APC framework that explicitly considers the effects of screening and menopausal hormone therapy (MHT). Breast cancer incidence in the presence of screening is computed by modeling the interaction between screening and the natural history of breast cancer.

Model S quantifies the effects of MHT in breast cancer incidence and mortality trends. For this purpose, the Stanford team developed a MHT dissemination model for women who were older than age 50 years before and after 2002. All parameters related to MHT modeling were calibrated to reproduce breast cancer incidence trends with increasing MHT use before 2002 and, as validation, to predict a rapid decline in incidence following the decline of MHT use in 2002. The Model S team tested the hypothesis that MHT increases tumor growth and decreases mammographic detectability and found that it is consistent with SEER data.

Tumor detection in the absence of screening is modeled as a stochastic process where the probability of detection increases over time, as a function of increasing tumor size. In the presence of screening, the tumor is either screen- or interval-detected (i.e., symptomatically detected in the interval between screening examinations). The impact of screening is computed by superimposing the times at screening on the natural history of the disease. If the tumor is in the preclinical phase and above the screen detection threshold at the time of screening, it is detected and its size is derived from the natural history model; if it is not detected by screening and becomes symptomatic before the next screening examination , it is classified as interval-detected. The dissemination of screening mammography is inferred from the common screening parameter with patterns of use derived from BCSC and national self-reported survey data (27). The detection function for screening mammography is characterized by a tumor size threshold;

above the threshold, the tumor is detectable, and below the threshold, the tumor is not detectable. Each individual is randomly assigned a screen-detection threshold, dependent on her age at the time of screening. Compared to older women, younger women have a higher detection threshold, which translates into lower program sensitivity.

Model S does not include DCIS and assumes that all invasive disease can progress to lethality. Treatment efficacy was originally modeled assuming proportional hazard; in other words, that the benefits were proportionally distributed across the years following diagnosis. In recent years, Model S has been updated to incorporate nonproportional hazard, to the effect that treatment depends on the ER status of each breast cancer case.

Wisconsin (Model W)

Model W is a discrete-event microsimulation model that uses a systems engineering approach to replicate breast cancer epidemiology in the U.S. population over time. It was developed at the University of Wisconsin and has been continuously maintained and enhanced for more than 10 years. Model W is a population-based model that simulates the lifetimes of individual women through the interaction of four main components: breast cancer natural history, detection, treatment, and mortality (5). Each woman enters the model at age 20 years and ages in 6-month cycles. Model W has several distinguishing features. First, the natural history component of the model incorporates heterogeneity in tumor characteristics. The model allows for a subset of clinically insignificant DCIS and early-stage invasive tumors that are more likely to be screen detected and do not lead to breast cancer mortality. On the other end of the spectrum, the model allows for another subset of tumors to be "aggressive" by entering them into regional and distant stages early in their development. These hypotheses about the natural history arose through the process of model calibration by testing the fit of alternative model structures to observed data. Using these natural history assumptions, Model W is able to closely replicate the dramatic increase in U.S. breast cancer incidence during the 1980s as well as the subsequent trends in breast cancer mortality. Second, the model incorporates second-order uncertainty in model outcomes. The calibration procedure uses acceptance sampling techniques whereby a joint posterior-like distribution across model parameters is produced. By conducting analyses with random samples from this distribution, the model results and conclusions reflect the effects of parameter uncertainty.

The interplay of the modules defines the individual life histories of the simulated women. Model output is highly customizable in its level of detail, like an omniscient cancer registry, and can include underlying disease states as well as observed clinical outcomes by age, race, and calendar year. The model uses common random numbers to reduce random variation in model outcomes and directly calculate quantities, such as overdiagnosis and lead-time (29). Model W incorporates second-order uncertainty in model outcomes such that model results and conclusions reflect the effects of parameter uncertainty. The model, programmed in C++, runs on both Microsoft Windows and UNIX platforms. A brief description of each component as currently programmed follows here.

Breast cancer occult inception is modeled as a function of a woman's birth year and age and accounts for secular trends in risk (23). After onset, cancers are assumed to grow according to a stochastic Gompertz-type model controlling for tumor size (30-32). Tumor spread is described by a Poisson process determined by tumor size and growth rate. Tumors are assigned a SEER historical stage (in situ, localized, regional, or distant) based on size and lymph node involvement at the time of diagnosis. The natural history component incorporates heterogeneity in tumor characteristics. A fraction of tumors are assumed to be of limited malignant potential (LMP) and do not pose a lethal threat. Model parameters that control the natural history and detection of breast cancer were estimated through the process of calibration to fit observed age-adjusted, stage-specific breast cancer incidence data from the SEER program and breast cancer mortality data from 1975 through 2010 reported by the National Center for Health Statistics. The calibration procedure uses acceptance sampling techniques whereby a joint posterior-like distribution across model parameters is produced (33). The model has been separately cross-validated against data from the Wisconsin Cancer Reporting System, a member of the North American Association of Central Cancer Registries, and the Iowa SEER registry. Separate analyses with this model produced results that were congruent with those from analyses based on other models of the natural history of breast cancer (5).

Breast cancer can be detected by one of two methods: by breast imaging or by symptoms (the combination of self-detection and clinical examination). Breast screening utilization, screening sensitivity, and the likelihood of symptomatic detection are functions of a woman's age and tumor size as well as calendar year to account for improvements in technology and increased awareness of the disease. The sensitivity of mammography has been calibrated to match observed estimates (34). Mammography utilization follows observed age-specific U.S. screening patterns (27), or the user can set screening utilization to follow fixed criteria, such as for starting and ending ages and frequency of use. Race-specific utilization of mammography was added to the model under recent funding (35, 36). The detection module can be configured for any screening test and has been extended to incorporate digital mammography (37).

Model W allows for a subset of clinically insignificant DCIS and early-stage invasive tumors that are more likely to be screen detected and do not lead to breast cancer mortality. As described above, a fraction of tumors are assumed to be of LMP and do not pose a lethal threat. LMP tumors are programmed with the following characteristics: 1) they start to grow at the same rate as lethal tumors; 2) they stop growing at a small size; and 3) they disappear if undetected after a fixed length of time.

The model assumes all women receive standard treatment at the time of detection. Adjuvant therapy with chemotherapy and/or endocrine therapy follows observed U.S. dissemination patterns (38, 39). Treatment effectiveness is a function of age at diagnosis, stage at diagnosis, ER/HER2 status, and receipt of adjuvant treatment and new treatment modalities, and is modeled independent of the method of cancer detection. An effective treatment is assumed to halt breast cancer progression ("cure"). Tumors that are not "cured" continue to grow until they reach a metastatic stage; survival time is assigned based on observed SEER cancer survival.

Analysis

For this modeling analysis, the six BWG models were used to assess screening outcomes for a cohort of women born in 1970 who are followed beginning at age 25 years (since breast cancer is rare before this age, accounting for only 0.08% of cases) until death. We report their screening outcomes from ages 40 to 100 years.

Model Data Input Parameters

The modeling groups began with a common set of age-specific variables for breast cancer incidence, breast density prevalence and probability of transition to lower breast density with age, digital mammography test characteristics, ER/HER2-specific treatments, and average and comorbidity-specific nonbreast cancer competing causes of death (**Table 1**). The models also used a common set of utility values.

In addition to the common parameters in **Table 1**, each model included model-specific inputs (or intermediate outputs) to represent preclinical detection times, lead-time, and age- and ER/HER2-specific stage distribution in screen- versus nonscreen-detected women on the basis of model structure (2-11). The model-specific parameters were based on reasonable assumptions about combinations of values that reproduce U.S. trends in incidence and mortality, including assumptions about proportions of DCIS that are nonprogressive and would not be detected without screening. Model W also assumed that some small invasive cancers are nonprogressive. Using a Bayesian approach, Model M accepted distributions of parameter sets that reproduce observed trends; these could include some nonprogressive DCIS and invasive cancer.

All models except one (Model M) used APC methods to project overall breast cancer incidence rates for the 1970 birth cohort in the absence of screening (13, 14, 23). Model M used incidence in the absence of screening based on rates reported to SEER from 1975 to 1979 without any specific cohort effects, so essentially a flat temporal trend.

To isolate the effect of technical screening effectiveness and assess the effect of screening on mortality while holding treatment constant, the models assumed 100 percent adherence to screening and receipt of and adherence to the most effective treatment.

Four groups used the age-specific digital mammography sensitivity values observed in the BCSC for detection of all cases of breast cancer (invasive and in situ) (**Table 2**).

Separate values were used for initial and subsequent mammography performed at either annual or biennial intervals, where annual interval was defined as 9 to 18 months between examinations and biennial interval as 19 to 30 months (unpublished data). One model (D) used these data directly as input variables (9) and three models (G-E, S, and W) used the data for calibration (5, 6, 8). The other models (E and M) used the BCSC data as a guide and fitted estimates from this and other sources (7, 10).

All women who had ER-positive invasive tumors received 5 years of hormonal therapy (tamoxifen if age at diagnosis was <50 years and anastrozole if ≥50 years) and nonhormonal treatment with an anthracycline-based regimen accompanied by a taxane. Women with ER-negative invasive tumors received nonhormonal treatment only. Those who had HER2-positive tumors received trastuzumab. Women with DCIS who had ER-positive tumors received hormonal therapy (39). Systemic treatment effectiveness was based on synthesis of recent clinical trials and was modeled as a proportionate reduction in mortality risk from the underlying ER/HER2-specific survival in the absence of therapy or the proportion cured (22), and assumed women received surgery (and radiation) as local therapy. To isolate the effectiveness of screening and to assess the impact of screening on mortality while holding treatment constant, models assumed 100 percent adherence to indicated treatment.

Benefits

Screening benefits occur because of reductions in breast cancer mortality due to detection at an earlier stage (or smaller size) than would have occurred without screening (with 100% treatment). Benefits are accumulated through ages 40 to 100 years to capture lifetime reductions in breast cancer mortality occurring years after the start (and end) of screening. We also examined breast cancer deaths averted, undiscounted life-years saved (LYS), and quality-adjusted life-years (QALYs) gained because of averted or delayed breast cancer death.

To quality adjust life-years, we applied a decrement in LYS related to age- and sex-specific average population general health (43, 44). Additional disutilities were then considered using multiplicative methods (44), including screening (-0.006 for 1 week), evaluation of a positive screen (-0.105 for 5 weeks), initial treatment by stage (for the first 2 years after diagnosis), and distant disease (for the last year of life for all women who die of breast cancer) (**Table 3**) (45, 46). The harm of overdiagnosis is captured by decrements in utility based on being in a cancer state and undergoing treatment but dying of other causes without a change in life expectancy.

Harms

As measures of the burden that a regular screening program imposes on a population, several different potential screening harms were examined: total number of mammograms, false-positive mammograms, benign biopsies, and overdiagnosis. We defined the rate of false-positive mammograms as the number of mammograms read as abnormal or needing further followup in women without cancer divided by the total number of screening mammograms. We defined benign biopsies post-hoc as the proportion of women with false-positive screening results who received a biopsy (47). We defined overdiagnosis as cases that would not have been clinically detected in a woman's lifetime in the absence of screening (because of lack of progressive potential or death from competing mortality).

Screening Scenarios

We compared model results for eight screening strategies varying by starting age (40, 45, or 50 years) and interval (annual, biennial, or hybrid [annual from ages 40 to 49 or 45 to 49 years and

biennial in the 50s]). All screening strategies stopped at age 74 years. We included a "no screening" base-case scenario to estimate the percent mortality reduction associated with any given strategy.

Strategies were arrayed from the least to the most harmful (or use of mammograms). We considered a strategy more efficient than a comparison strategy if it resulted in more benefits for a given increase in harm (or use). A strategy that entails more harms (or use) but less benefit was considered inefficient or dominated by other strategies. After eliminating all dominated and weakly dominated strategies, we represented the remaining strategies as points on a graph plotting the average amount of harm versus benefit for each model. We obtained the efficiency frontier for each graph by identifying the sequence of points that represented the largest incremental gain in percentage of mortality reduction (or LYS or QALYs) per additional harm entailed. Screening strategies that fall on this frontier were considered the most efficient (i.e., no alternative exists that provides more benefit with fewer harms).

Subpopulation and Sensitivity Analysis

We conducted additional analyses to evaluate whether the ranking of strategies changes when different subpopulations or input variable levels are modeled. First, we investigated the effect of breast density on harms and benefits. We considered four density subpopulation groups for analyses (a=entirely fatty; b=scattered density; c=heterogeneously dense; and d=extremely dense) and grouped them for reporting into low density (a and b) and high density (c and d). Density was assigned at age 40 years and could decrease by one level or remain the same at age 50 years and again at age 65 years based on observed age-specific prevalence rates from the BCSC (unpublished data) (**Table 4**). Density modifies mammography sensitivity and specificity based on age-, interval-, and screening round–specific data observed in the BCSC (unpublished data). Density also modifies age-specific risk of developing breast cancer (age groups, 40 to 49, 50 to 64, and ≥65 years), using marginal population density–related risk in each age group as the referent group (BCSC personal communication).

We also increased the underlying risk of breast cancer from nondensity-related causes from a relative risk of 1 (average population) to test the impact of risk level on benefits and harms for women in their 40s. Risk levels were applied at age 40 years and were assumed to persist at the same levels for the remainder of a woman's life. The exemplar risk levels were 1.3 (e.g., corresponding to postmenopausal obesity) (48), 2 (e.g., family history of one first-degree relative with onset before age 50 years) (49), and 4 (e.g., family history of ≥2 first-degree relatives) (49, 50). Higher risk levels, such as having a *BRCA1* or *BRCA2* gene mutation, were not considered since such women have specific screening guidelines. Finally, we examined the impact of combinations of breast density and risk level (**Table 5**). We made the simplifying assumption that risk only affects incidence (except breast density, which also affects sensitivity and specificity), and did not modify other aspects of disease.

In other subpopulation analyses, we examined the impact of comorbidity-specific versus overall population competing mortality on upper ages of screening cessation based on comorbidity-specific life expectancy (51-53). We compared results for continuing to screen biennially past age 74 years among women with lower than average comorbidity or stopping earlier than age 74

years among those with moderate or high comorbidity. These analyses only included women who have survived and not developed breast cancer up until the point where screening is to be extended or stopped.

In sensitivity analysis, we considered the impact of varying the values for disutilities to identify if there was a level of disutility where there were no longer meaningful screening benefits.

Model Validation and Uncertainty

Each model had a different structure and assumptions and some varying input variables, so no single method could 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, mortality, and stage distribution to those reported by the SEER program for the period of 1975 to 2010. In our previous work, results of each model accurately projected incidence and mortality trends by ER status for the period of 1975 to 2000 (4). Next, we approximated the AGE screening trial (54), assuming perfect adherence to invitation for annual screening and 13-year followup of women ages 40 to 49 years. In addition, one model also estimated AGE trial effects assuming actual trial attendance patterns (54). Last, we examined consistency of results across models. Overall, using six models to project a range of plausible screening outcomes provided implicit cross-validation, with the range of results from the models as a measure of uncertainty.

Chapter 3. Results

Validation

The models closely estimated observed U.S. trends in incidence and mortality, stage distribution (not shown), and the AGE trial results. Using input for actual dissemination of screening and treatment in the United States, as shown in **Figures 2a–c**, the models all captured the major trends in incidence over time, with early increases with the advent of mammography in the mid-1980s, a downturn in the 2000s, and then leveling off. The models also captured the general shape of mortality decreases over time.

The models also simulated the conditions of the AGE trial, a screening trial conducted in the United Kingdom to evaluate the benefits of screening women in their 40s. Women were randomized and invited to seven to eight rounds of annual screening versus no screening; all women were invited to biennial screening at age 50 years, and breast cancer mortality was tracked over 13 years. Screening was done with two views initially and a single view thereafter, and included plain film and digital mammography. Based on an intention-to-treat analysis of women invited to screen, the trial found a 0.83 (95% confidence interval [CI], 0.66 to 1.04) relative risk of death from breast cancer in the group invited to screen versus controls. In analyses that examined actual screening use in the trial (i.e., not all women invited to screening actually attended), the relative risk reduction was 0.76 (95% CI, 0.51 to 1.01). The models considered actual use but assumed 100 percent compliance with screening and 100 percent use of treatment based on U.S. approaches. Thus, the models projected a relative risk of 0.72 (95% CI, 0.65 to 0.75), a similar but slightly more optimistic risk reduction for screening than that of women in their 40s (**Table 6**). However, these results indicate good validation by models of trial results.

Probability of Disease

In an unscreened population, the models predicted a median 12.9 percent cumulative probability of developing breast cancer from ages 40 to 100 years (range across models, 12.0% to 14.0%). Without screening, the median probability of dying of breast cancer was 2.5% (range, 1.5% to 3.2%). Thus, if a screening strategy leads to a 30 percent reduction in breast cancer mortality, the probability of breast cancer death would be reduced from 2.5 to 1.8 percent, or 7.5 deaths averted per 1,000 women screened.

Benefits

The models produced consistent rankings of screening strategies (**Tables 7a–c**). Biennial screening from ages 50 to 74 years yielded a median 25.8 percent reduction in lifetime breast cancer mortality compared to no screening (range across models, 24.1% to 31.8%). This translates into 7.1 (range, 3.8 to 8.7) deaths averted per 1,000 women (**Table 8**). Annual screening from ages 50 to 74 years was dominated in half of the models, but annual screening

from ages 40 to 74 years could result in a 37.8 percent median mortality reduction (range, 32.5% to 43.6%) and avert 10.1 (range, 6.3 to 11.2) deaths compared to no screening. Biennial strategies maintained, on average, 81.2 percent (mortality reduction range across strategies and models, 68.3% to 98.9%) of the benefit of annual screening (**Table 9**). Rankings were similar for LYS and QALYs, except that annual screening from ages 50 to 74 years was dominated in all six models using these benefit metrics (**Tables 7b** and **c**). **Table 8** summarizes the median benefits and range across models for all strategies.

Incremental Benefits

Incremental benefits of starting screening at age 40 versus 50 years were slightly greater in terms of breast cancer deaths averted for annual versus biennial screening (median, 1.3 [range, 1.1 to 1.7] vs. 1.0 [range, 0.8 to 1.7] per 1,000 women, respectively). Screening starting at age 45 years (vs. age 50 years) resulted in a similar pattern, but about 0.6 fewer deaths averted per 1,000 women than when starting at age 40 years (**Tables 10a** and **b**).

Harms

All of the models projected that there are nearly twice as many false-positive results and more overdiagnosed cases with annual versus biennial schedules and starting at age 40 versus 50 years (**Table 11**). For instance, if biennial screening begins at age 40 years instead of age 50 years, for every 1,000 women screened there would be a median 1 more death averted, but 576 more false-positive results, 67 more induced biopsies among women with false-positives, and 2 additional overdiagnosed cases. The most intensive screening strategy (annual screening from ages 40 to 74 years) would save 3 more deaths than biennial screening from ages 50 to 74 years, with nearly 2,000 more false-positives and 11 more overdiagnosed cases. The majority of these overdiagnosed cases are DCIS. A hybrid strategy (annual screening from ages 45 to 49 years and biennial screening from ages 50 to 74 years) preserves the majority of the gains accrued with annual screening beginning at age 40 years, with less harm.

If screening occurs from ages 50 to 74 years, annual screening averts a median 2 more deaths from breast cancer for every 1,000 women screened, but nearly doubles the number of false-positives and increases overdiagnosed cases (**Table 11**).

The percent of all cases diagnosed that are overdiagnosed increases as the age of initiation decreases from 50 to 40 years, and increases the most with annual versus biennial screening (**Table 12**).

Efficiency Frontiers

Biennial strategies starting at ages 40, 45, and 50 years would generally be considered efficient in most models when examining the balance of screening benefits and harms across most metrics (**Figures 3** and **4**). The hybrid strategy of annual screening starting at age 40 or 45 years was also efficient. The most intensive strategy (annual screening from ages 40 to 74 years) was the upper

anchor in all efficiency assessments. Annual screening from ages 50 to 74 years was dominated and not on the efficiency frontier in almost all the models (e.g., 4/6 models for mortality reductions per mammogram, or 6/6 for LYS per mammogram) and for nearly all outcome metrics (two of which are shown here). Figures for QALYs, false-positive mammograms, and overdiagnosed cases are not shown, but patterns were similar as for other metrics, and all models indicated that annual screening from ages 50 to 74 years was dominated for these metrics.

Subpopulation Analyses

The ranking of strategies did not change when screening occurred in groups of women with increasing risk levels (from 1.3 to 4 times the average); annual screening from ages 50 to 74 years remained dominated by other approaches across risk levels (**Table 13a**). However, the balance of benefits and harms varied by risk level. As risk increases, benefits increase and harms generally decrease. For example, among women with a 2- to 4-fold increase in risk, annual screening starting at age 40 years (or 45 years) would have a similar or more favorable harm to benefit ratio (based on false-positives or overdiagnosed cases) than biennial screening in average-risk women from ages 50 to 74 years (**Table 13b**). For women with even a 1.3-fold increase in risk, biennial screening starting at age 40 years would have a similar or more favorable harm to benefit ratio as biennial screening in average-risk women from ages 50 to 74 years (**Table 13b**).

Breast density varies with age, but considering breast density did not affect the ranking of strategies versus no screening; annual screening from ages 50 to 74 years remained dominated for all breast density groups (**Table 13c**). Women with low breast density (BI-RADS density a and b) had more of their cases detected and therefore greater mortality reduction (vs. no screening) than those with higher breast density (BI-RADS density c and d) (**Table 13c**). However, when women with dense breasts had their cancer detected, there were more deaths averted and greater LYS compared to women with less dense breasts (**Table 13c**), since there were more cases in women with high breast density, and high breast density is most often seen in younger women with longer life expectancy. The ratios of harm to benefit are summarized in **Table 13d**. The aforementioned trends are also consistently seen when screening strategies consider both risk and density (**Tables 13e and 13f**).

For women with no comorbidity who have an average of a 17-year remaining life expectancy (**Table 14**), screening would be efficient through age 78 to 80 years and would have a minimal increase in overdiagnosis compared to stopping at age 74 years. However, for women with severe comorbidity, screening cessation could occur at about age 68 years (**Figure 5**).

Sensitivity Analyses

Consideration of utilities decreased estimates of LYS (**Table 15**), but did not affect the ranking of strategies (**Tables 7a–c**). The largest decrement related to quality adjustment accrued because of declines in general health as women aged. There were also substantial percent decrements in QALYs attributed to the disutility of undergoing diagnostic evaluation of an abnormal screening

examination and having cancer (approximate 15% decrease for each QALY beyond general health). The disutility associated with undergoing screening itself had a minimal impact on QALYs. There were proportionally fewer losses in QALYs in biennial versus annual screening strategies, given that the latter had a greater number of false-positives relative to benefits than biennial screening. Overall, there were persistent QALY gains under all screening strategies, but the magnitude became smaller when the highest disutility estimates were used (**Figure 6**; **Table 15**).

Chapter 4. Discussion

Limitations

The conclusions about the ranking of screening strategies are fairly robust across six independent models and should provide greater credibility than inferences based on one model alone. Despite the general consistency of results across models, there are several caveats that should be considered in evaluating the modeling results. First, our goal was to consider and compare program efficacy, so we assumed 100 percent use of periodic screening, prompt evaluation of abnormal results, and full use of optimal treatment. Actual benefits will always fall short of the projected results since use and access is not perfect. If actual usage patterns vary systematically by age, risk, or other factors, it is possible that the ranking of strategies could change.

Second, there is expected variability across models in absolute estimates of benefit and harm metrics based on differences in model structure and assumptions. However, the value of having multiple models collaborate using the same inputs is that they produce a credible range of results. The overall conclusions were also similar across models for mostly all analyses. The models have the greatest variability in results for overdiagnosis since this is an unobservable event for which there is no primary biologically-based data. Thus, overdiagnosis must be inferred indirectly. Many methods for this have been proposed, and there is no gold standard approach (55, 56). Modeling makes a useful contribution to estimating overdiagnosis since it explicitly considers lead-time and competing mortality and takes a lifetime perspective. Overall, using multiple models produces a range of results for overdiagnosis (and other screening outcomes) that can be useful to decisionmakers.

Third, we did not consider other imaging technologies, such as computer-assisted diagnosis (57), tomosynthesis, or magnetic resonance imaging, or the risk of breast cancer induced by radiation related to screening. There are emerging data on their performance in general populations (58), so this will be important to consider once additional data are available. Next, we assumed that risk factors influenced the incidence but not natural history of breast cancer, and that, except for breast density–related risk, this risk was constant over a woman's lifetime. However, certain risk factors, such as family history, are age-dependent in their effects (49, 59). Thus, our estimates of benefit could be overestimated or underestimated based on specific risk factors (16). Additional modeling of risk and breast density and consideration of a wider array of screening approaches is warranted.

While several countries are now testing genetic risk–based screening, we did not consider polygenic risk related to panels of gene polymorphisms. This is an important emerging area for future research, especially as genetic profiles can be linked to risk of specific molecular subtypes (60-62). Compared to our earlier research (3), the models all estimated similar but somewhat greater mortality reductions from screening (e.g., 22% vs. 25.8% median mortality reduction with biennial screening from ages 50 to 74 years in 2009 vs. current models, respectively). The primary reasons for this modeled improvement relate to the increased sensitivity of digital versus film mammography, advances in molecular subtype–directed adjuvant therapy, and changes in underlying breast cancer trends.

Additionally, while we externally validated the models to the AGE trial, we did not match all of the trial conditions. It will be important to conduct a fuller model validation of the trial in the future. It is also unclear which of the different outcome metrics is optimal for use in decisionmaking. However, the conclusions about which screening strategies are on the efficiency frontier were fairly consistent across metrics. Finally, these analyses were designed to provide modeling data for use in public health decisionmaking for populations of women; the results are not intended to guide individual screening decisions.

Summary

This report summarizes research by six established models with differing approaches and assumptions to estimate the potential efficacy of various screening strategies for the U.S. population and subpopulation groups based on risk, breast density, and comorbidity. All six modeling groups projected some benefits associated with screening women from ages 40 to 49 years. The models consistently ranked strategies and concluded that biennial screening strategies were most frequently on the efficiency frontier. Screening initiation at age 40 years had the greatest benefit but also the greatest harms. Thus, decisions depend on tolerance for additional false-positive screens, biopsies, and overdiagnosed cases. The ranking of strategies was not affected by risk level or breast density, but absolute benefts and harms did change by risk level and breast density. Annual screening in women ages 40 to 74 years with a 2- to 4-fold increased risk or biennial screening in those with a 1.3-fold increased risk have a comparable ratio of benefits to harms as biennial screening in the general population from ages 50 to 74 years. Consideration of disutilities reduced but did not eliminate the magnitude of benefit from all strategies. Among women with severe or moderate levels of comorbidity, harms of screening seemed to outweigh benefits prior to age 74 years. Choices about optimal ages of initiation and cessation will ultimately depend on program goals, weight attached to the balance of harms and benefits, and considerations of efficiency.

References

1. Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR, Hadorn D. Assessing the effectiveness of health interventions for cost-effectiveness analysis. In: Gold MR, Seigel J, Russell LB, Weinstein MC (eds). Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996.
2. 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):1784-92.
3. 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 Intern Med*. 2009;151(10):738-47.
4. 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;106(11).
5. 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.
6. 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.
7. 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):30-6.
8. 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.
9. Lee S, Zelen M. A stochastic model for predicting the mortality of breast cancer. *J Natl Cancer Inst Monogr*. 2006;(36):79-86.
10. 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.
11. 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.
12. Cronin KA, Feuer EJ, Clarke LD, Plevritis SK. Impact of adjuvant therapy and mammography on U.S. mortality from 1975 to 2000: comparison of mortality results from the CISNET breast cancer base case analysis. *J Natl Cancer Inst Monogr*. 2006;(36):112-21.
13. Gangnon RE, Sprague B, Stout N, Alagoz O, Trentham-Dietz A. Age-period-cohort analysis of invasive breast cancer incidence in the United States, 1935-2008 [Abstract]. *Am J Epidemiol*. 2012;175(Suppl 11):S1-145.

14. 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. *Cancer Epidemiol Biomarkers Prev*. 2015; Mar 18.
15. 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;136:823-35.
16. 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):609-17.
17. van Ravesteyn NT, Schechter CB, Near AM, Heijnsdijk EA, Stoto MA, Draisma G, et al. Race-specific impact of natural history, mammography screening, and adjuvant treatment on breast cancer mortality rates in the United States. *Epidemiol Biomarkers Prev*. 2011;20(1):112-22.
18. Mandelblatt J, van Ravesteyn N, Schechter C, Chang Y, Huang AT, Near AM et al. Which strategies reduce breast cancer mortality most? Collaborative modeling of optimal screening, treatment, and obesity prevention. *Cancer*. 2013;119(14):2541-8.
19. Lee SJ, Zelen M. Mortality modeling of early detection programs. *Biometrics*. 2008;64(2):386-95.
20. Lee SJ, Zelen M. Modelling the early detection of breast cancer. *Ann Oncol*. 2003;14(8):1199-202.
21. Lee SJ, Zelen M. Scheduling periodic examinations for the early detection of disease: applications to breast cancer. *J Am Stat Assoc*. 1998;93(4444):1271-81.
22. 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):432-44.
23. Holford TR, Cronin KA, Mariotto AB, Feuer EJ. Changing patterns in breast cancer incidence trends. *J Natl Cancer Inst Monogr*. 2006;(36):19-25.
24. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am*. 1992;30(1):187-210.
25. 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):625-51.
26. National Cancer Institute. Surveillance, Epidemiology and End Results Program (SEER): SEER*Stat Software. Accessed at http://seer.cancer.gov/seerstat/ on 7 April 2015.
27. Cronin KA, Yu B, Krapcho M, Miglioretti DL, Fay MP, Izmirlian G, et al. Modeling the dissemination of mammography in the United States. *Cancer Causes Control*. 2005;16(6):701-12.
28. Plevritis SK, Salzman P, Sigal BM, Glynn PW. A natural history model of stage progression applied to breast cancer. *Stat Med*. 2007;26(3):581-95.
29. Stout NK, Goldie SJ. Keeping the noise down: common random numbers for disease simulation modeling. *Health Care Manag Sci*. 2008;11(4):399-406.

30. Shwartz M. An analysis of the benefits of serial screening for breast cancer based upon a mathematical model of the disease. *Cancer*. 1978;41(4):1550-64.
31. Shwartz M. Validation and use of a mathematical model to estimate the benefits of screening younger women for breast cancer. *Cancer Detect Prev*. 1981;4(1-4):595-601.
32. Spratt JA, von Fournier D, Spratt JS, Weber EE. Decelerating growth and human breast cancer. *Cancer*. 1993;71(6):2013-9.
33. Stout NK, Knudsen AB, Kong CY, McMahon PM, Gazelle GS. Calibration methods used in cancer simulation models and suggested reporting guidelines. *Pharmacoeconomics*. 2009;27(7):533-45.
34. Heine JJ, Malhotra P. Mammographic tissue, breast cancer risk, serial image analysis, and digital mammography, 1: tissue and related risk factors. *Acad Radiol*. 2002;9(3):298-316.
35. Batina NG, Trentham-Dietz A, Gangnon RE, Sprague BL, Rosenberg MA, Stout NK, et al. Variation in tumor natural history contributes to racial disparities in breast cancer stage at diagnosis. *Breast Cancer Res Treat*. 2013;138(2):519-28.
36. Using computer simulation to investigate racial disparities in the natural history of breast cancer. Presented at the 31st Annual Meeting of the Society for Medical Decision Making, October 18-21, 2009, Hollywood, CA.
37. Tosteson AN, Stout NK, Fryback DG, Acharyya S, Herman BA, Hannah LG et al. Cost-effectiveness of digital mammography breast cancer screening. *Ann Intern Med*. 2008;148(1):1-10.
38. Mariotto AB, Feuer EJ, Harlan LC, Abrams J. Dissemination of adjuvant multiagent chemotherapy and tamoxifen for breast cancer in the United States using estrogen receptor information: 1975-1999. *J Natl Cancer Inst Monogr*. 2006;(36):7-15.
39. National Comprehensive Cancer Network. Clinical Practice Guidelines in Oncology: Breast Cancer. 2014.
40. Carter SB, Gartner SS, Haines MR, et al. Historical Statistics of the United States, Volume One: Population. New York: Cambridge University Press; 2006.
41. Cronin KA, Mariotto AB, Clarke LD, Feuer EJ. Additional common inputs for analyzing impact of adjuvant therapy and mammography on U.S. mortality. *J Natl Cancer Inst Monogr*. 2006;(36):26-9.
42. 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):667-76.
43. 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.
44. Fryback DG, Lawrence WF. Dollars may not buy as many QALYs as we think: a problem with defining quality-of-life adjustments. *Med Decis Making*. 1997;17(3):276-84.
45. 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):774-82.
46. 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;49:538-44.

47. 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.
48. 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):1724-33.
49. 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):635-48.
50. 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):1204-14.
51. 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):1376-85.
52. 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.
53. Dewey ME, Saz P. Dementia, cognitive impairment and mortality in persons aged 65 and over living in the community: a systematic review of the literature. *Int J Geriatr Psychiatry*. 2001;16(8):751-61.
54. 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):2053-60.
55. Etzioni R, Gulati R. Oversimplifying overdiagnosis. *J Gen Intern Med*. 2014;29(9):1218-20.
56. 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):831-8.
57. 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):580-7.
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):2499-507.
59. 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):165-75.
60. 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.
61. Stevens KN, Vachon CM, Couch FJ. Genetic susceptibility to triple-negative breast cancer. *Cancer Res*. 2013;73(7):2025-30.
62. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013;45(4):353-2.

Figure 1. Natural History of Breast Cancer in Model D

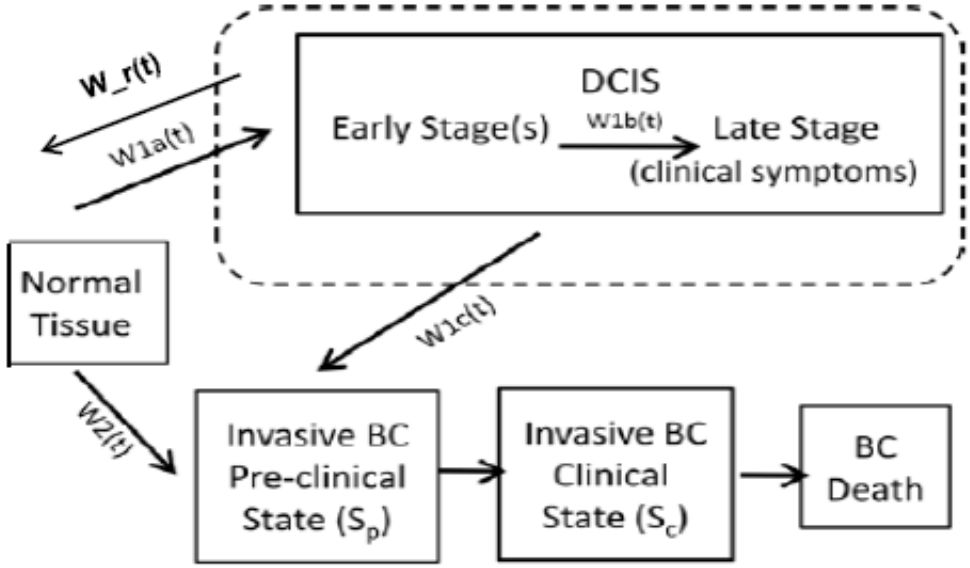


Figure 2a. Modeled Invasive Incidence Using Actual U.S. Screening vs. SEER, Ages 40 to 100 Years

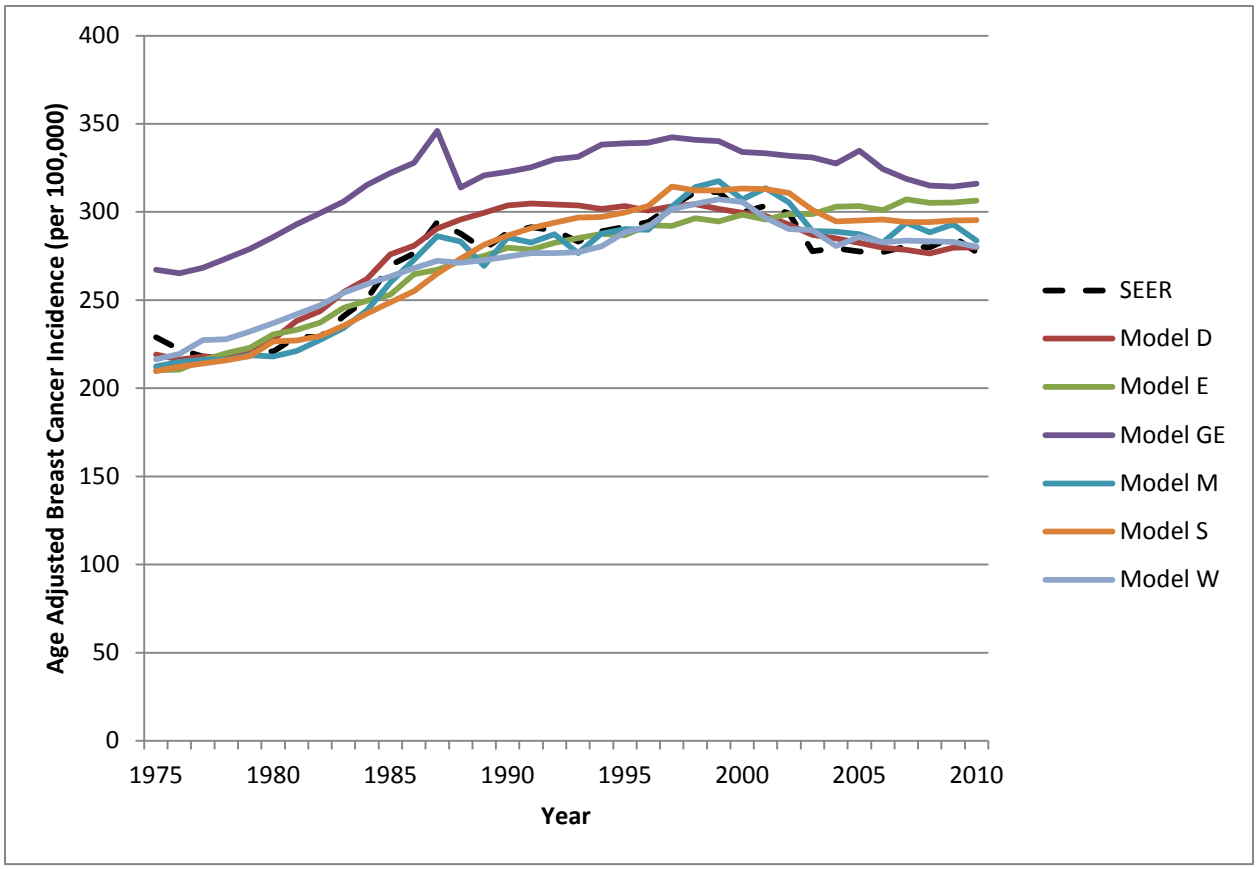
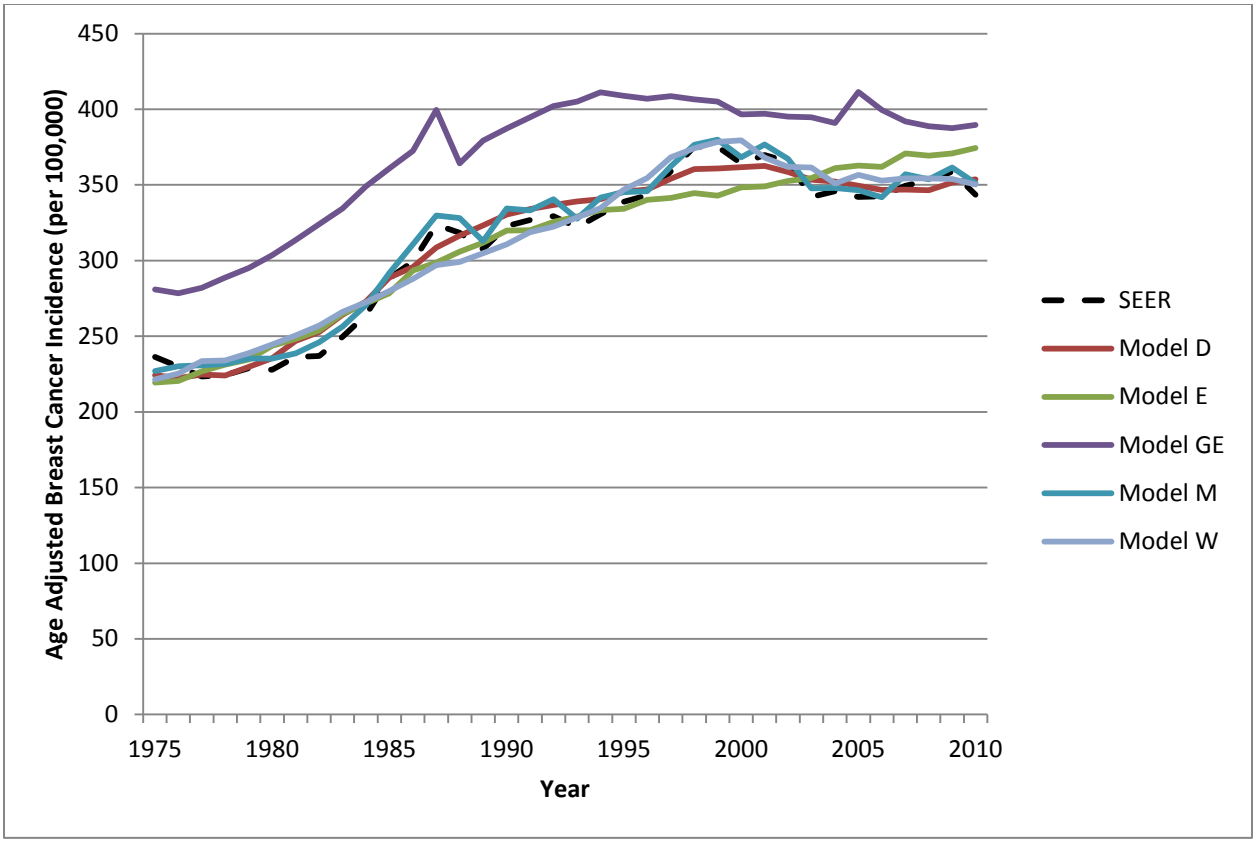


Figure 2b. Modeled Invasive and DCIS* Incidence Using Actual U.S. Screening vs. SEER, Ages 40 to 100 Years



*Model S does not include DCIS; Model G-E peaks in 2005 due to transition from plain film to digital mammography in a single year, since digital mammography has higher sensitivity/DCIS detection than plain film mammography.

Figure 2c. Modeled Mortality Using Actual U.S. Screening and Treatment vs. Observed, Ages 40 to 100 Years

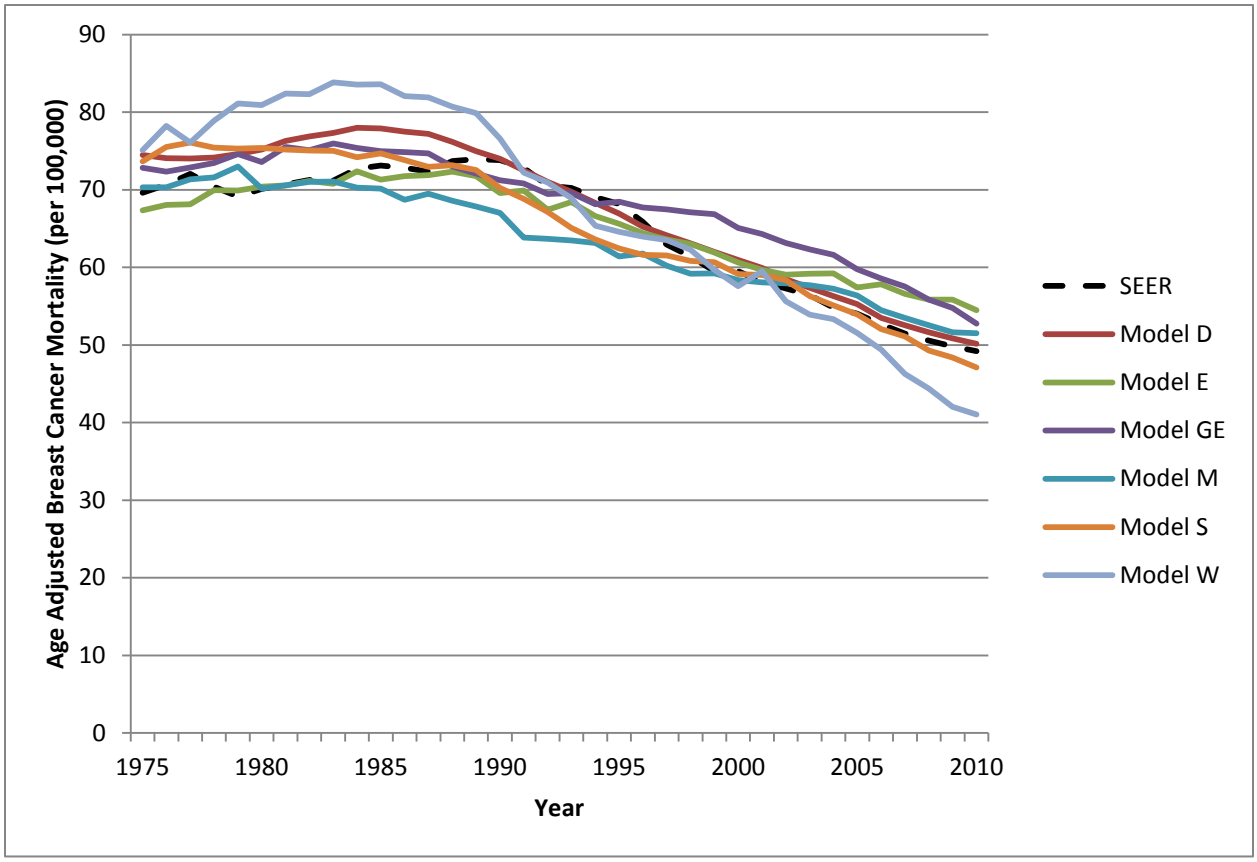
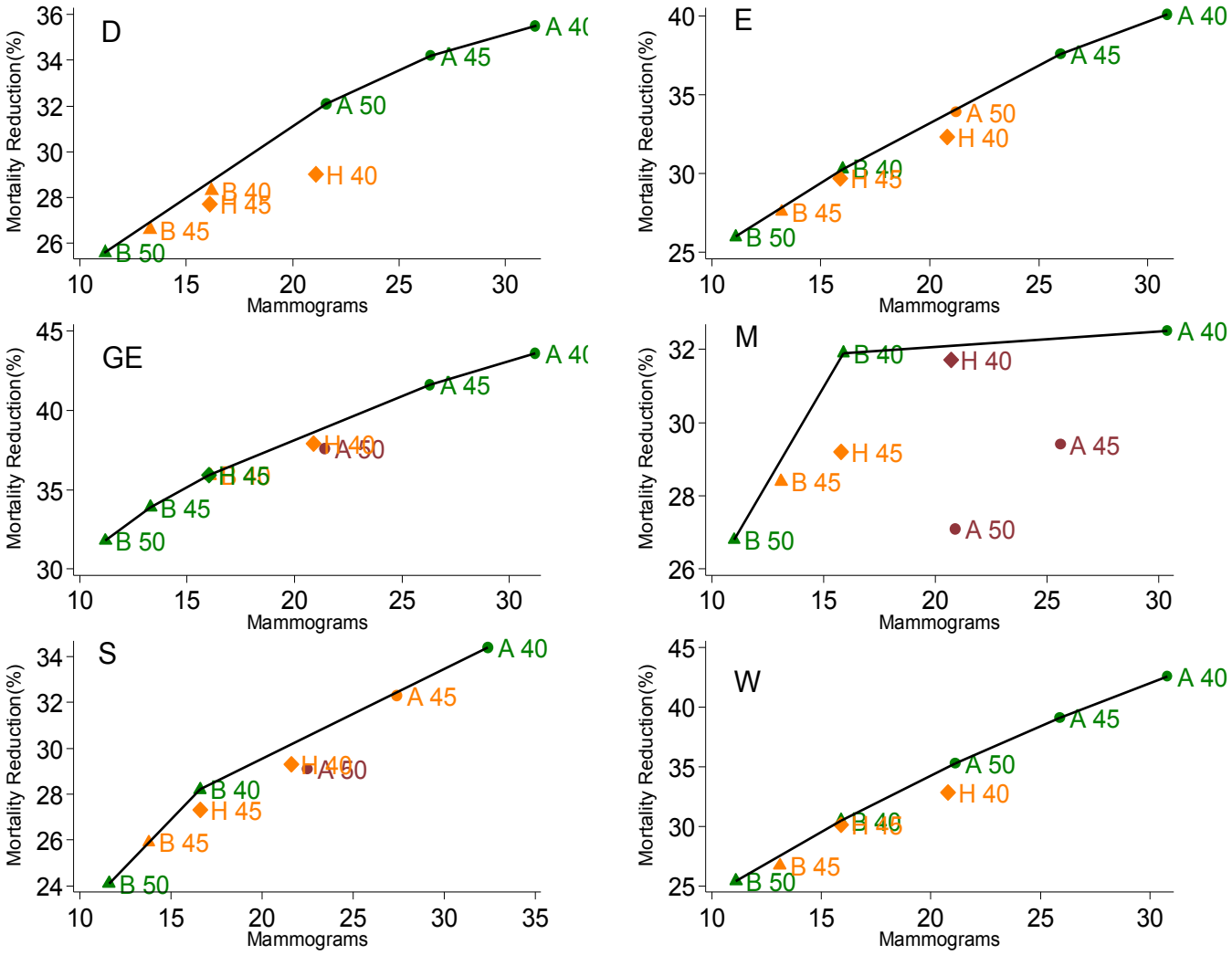
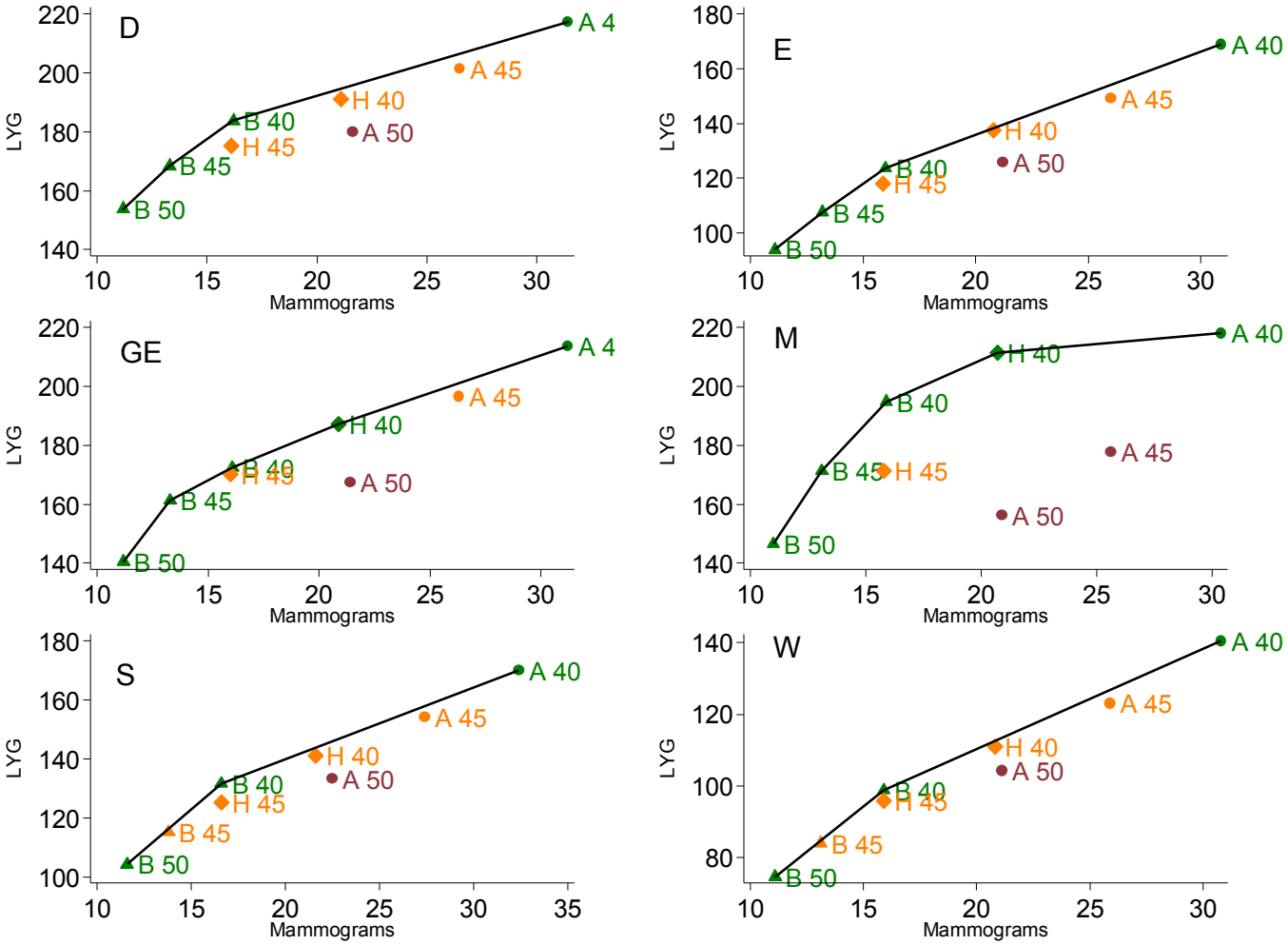


Figure 3. Efficiency Frontier for Harms (Average Number of Screening Examinations) and Benefits (Percent Mortality Reduction) by Model and Screening Strategy



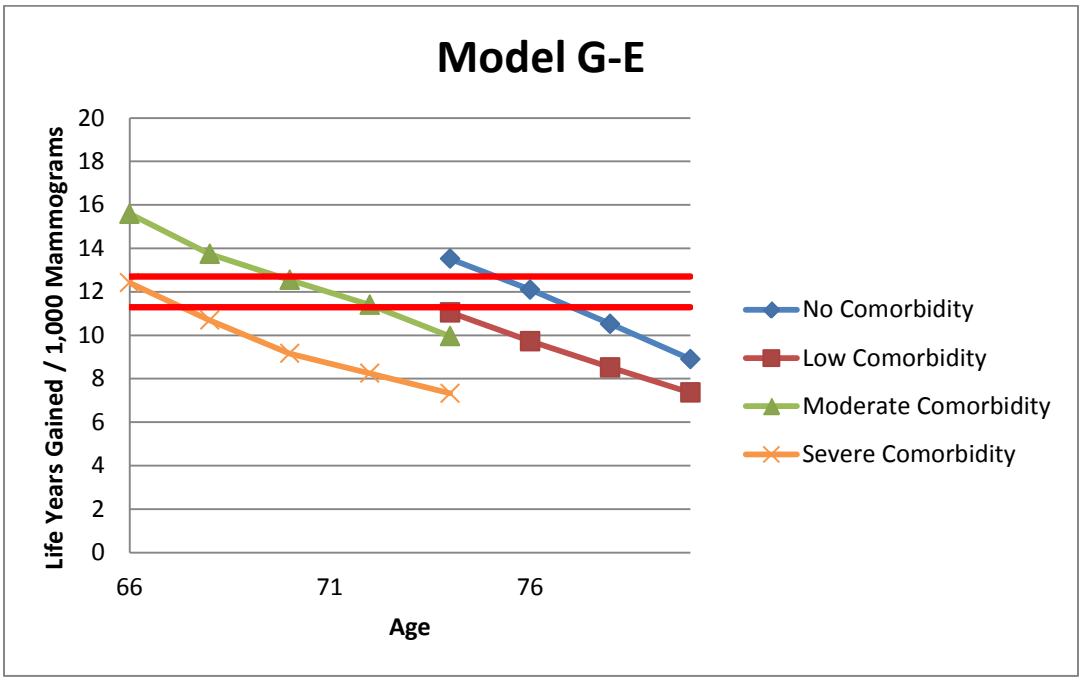
The panels show an efficiency frontier graph for each model. The graph plots the average number of mammograms performed per 1,000 women against the percentage of mortality reduction for each screening strategy (vs. no screening). We plot efficient strategies (i.e., those in which increases in use of mammography results in greater mortality reduction than the next least-intensive strategy) in all six models. The line between strategies represents the "efficiency frontier." Strategies on this line would be considered efficient because they achieve the greatest gain per mammography performed compared with the point (or strategy) immediately below it. Points that fall below the line are not considered as efficient as those on the line. When the slope in the efficiency frontier plot levels off, the additional reductions in mortality per unit increase in use of mammography are small relative to the previous strategies and could indicate a point at which additional investment (use of screening) might be considered as having a low return (benefit). Green strategies are efficient; yellow strategies are close to the efficiency frontier; and red strategies are dominated (inefficient).

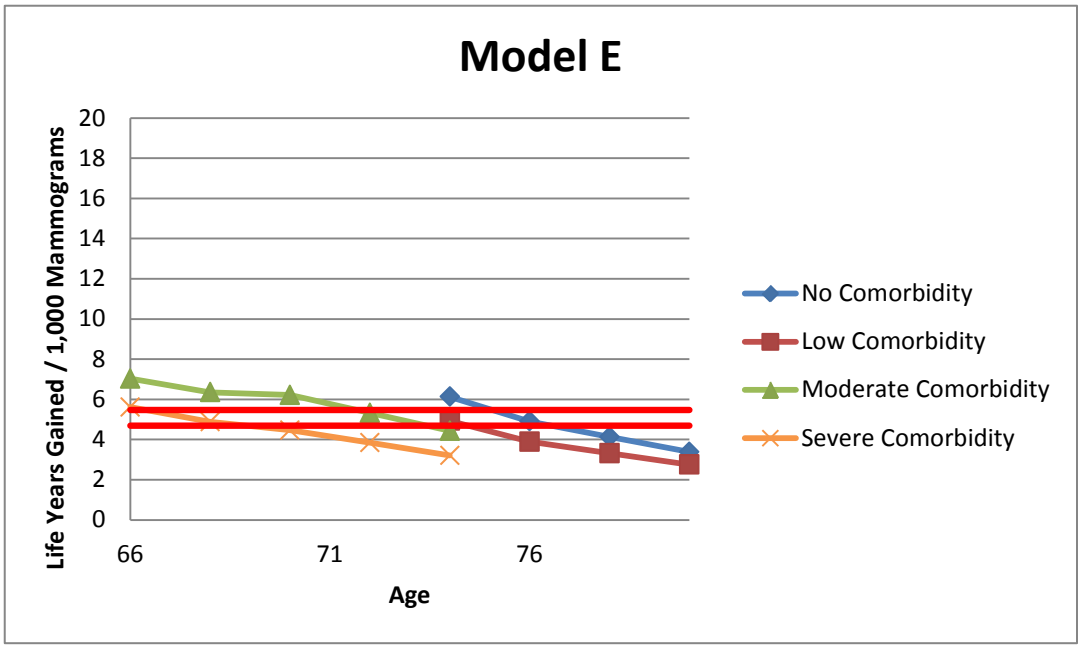
Figure 4. Efficiency Frontier for Harms (Average Number of Screening Examinations) and Benefits (Life-Years Gained) by Model and Screening Strategy



The panels show an efficiency frontier graph for each model. The graph plots the average number of mammograms per 1,000 women against the life-years gained for each screening strategy (vs. no screening). We plot efficient strategies (i.e., those in which increases in mammography use results in greater life-years gained than the next least-intensive strategy) in all six models. The line between strategies represents the "efficiency frontier." Strategies on this line would be considered efficient because they achieve the greatest gain per mammography compared with the point (or strategy) immediately below it. Points that fall below the line are not as efficient as those on the line. When the slope in the efficiency frontier plot levels off, 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 (benefit). Green strategies are efficient; yellow strategies are close to the efficiency frontier; and red strategies are dominated (inefficient).

Figure 5. Incremental Benefit to Harm Ratio of an Additional Biennial Screening Mammogram Relative to Stopping Screening at Age 74 Years by Comorbidity Level–Specific Life Expectancy and Model: Life-Years Gained per 1,000 Mammograms





Results are for having 1 additional mammogram after the last one at age X. The solid red lines represent the results in terms of life-years gained per 1,000 women for screening once more at age 76 years vs. stopping at age 74 years for a population with average comorbidity level–associated life expectancy. Results within this "threshold" are those implied by recommendations to stop screening at age 74 years. The life-years gained per 1,000 biennial screens are shown at each age for four comorbidity groups: none (blue), low (dark orange), moderate (green), and severe (yellow). These analyses start with populations of women who are alive and have never developed breast cancer prior to the age of the start of the simulation. For example, in both models, women with no comorbidity who are alive and cancer-free at age 74 years are screened again at age 76 years. Since this result is within current thresholds, those who are still alive and cancer free at age 76 years might consider another screen at age 78 years and remain within the threshold. Conversely, those with severe comorbidity (yellow line) who are alive and cancer free at age 66 years will fall below the threshold at age 68 years, indicating that screening might stop earlier.

Figure 6. Quality of Life Adjustments by Health State Across Strategies

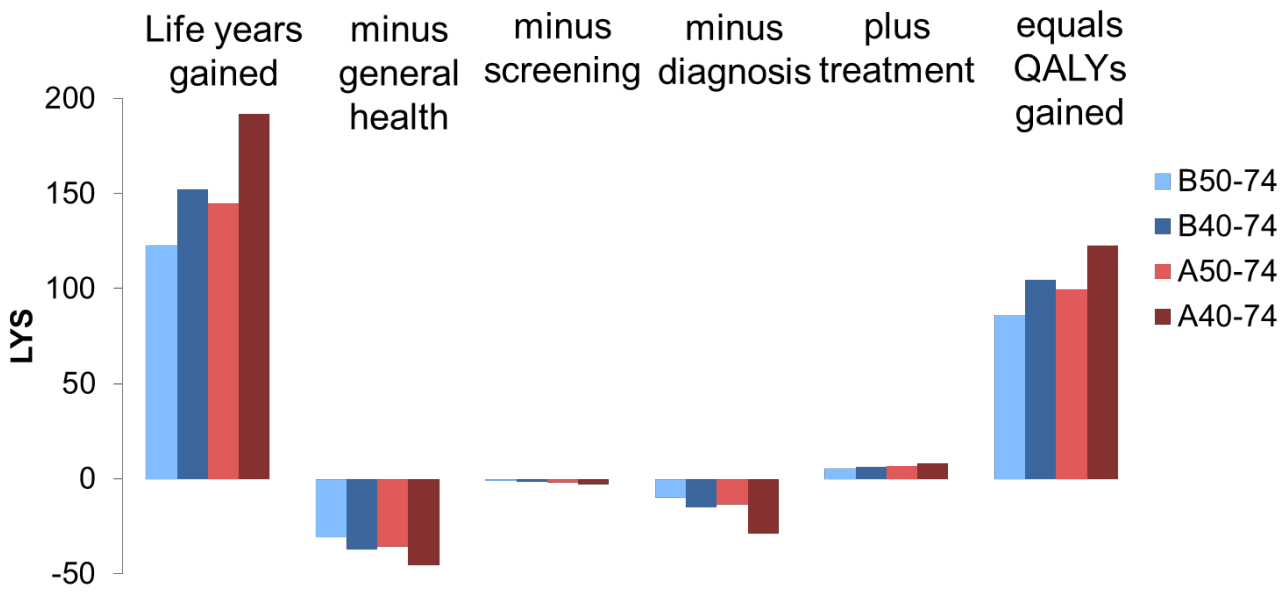


Table 1. General Model Data Input Parameters and Model Assumptions

Parameter	Description	Data Source
Demographic		
Birth cohorts	1970 birth cohort	(40)
Natural History		
Incidence in the absence of screening	An age-period-cohort model is used as a starting point for most models (except Model M)	(14)
Stage distribution	Stage distribution among clinically-detected and digital screen-detected women by age group (<50, 50–64, ≥65 years), screening round (first, subsequent), and screening interval (annual, biennial). *Screen-detected:* Cancer diagnosed within 12 months after a positive screen and prior to the next screening mammogram (with and without self-reported symptoms). *Interval-detected:* Cancer diagnosed within 6 months after or 30 days before a diagnostic mammogram, with a screening mammogram within 42 months prior to that mammogram. *Clinical-detected:* A diagnostic mammogram between 6 months prior to and 30 days after the cancer diagnosis and no prior mammogram within 3.5 years (42 months) of the diagnostic mammogram.	BCSC data from 1994–2013 (digital from 2003–2013; excluded first year of transition to digital)
ER/HER2 joint distribution	The probability of ER/HER2 conditional on age and stage at diagnosis	BCSC
Sojourn time	Sojourn time by joint ER/HER2 status and age group	(4)
Mean stage dwell time/tumor growth rates	Varies by models based on model structure; can vary by age and/or ER/HER2 status	(6, 10)
Other-cause mortality	Age- and cohort-specific all-cause mortality rates by year	Jackson H et al (personal communication)
Screening		
Mammography rates	Assume all women age ≥40 years are screened by digital mammography	(27, 41)
Sensitivity/ detection rates of digital screening	Sensitivity of initial and subsequent digital mammography by age group, screening interval (annual, biennial), and breast density	BCSC
Specificity	False-positive mammograms are calculated as the difference between the overall number of positive mammograms in a screening scenario minus the number of positive mammograms among breast cancer cases	
Survival	26-year breast cancer survival before adjuvant treatment by joint ER/HER2 status, age group, and AJCC/SEER stage or tumor size	(4)
Treatment		
Treatment patterns	Assume receipt of and adherence to the most effective available treatment	1997–2010 (38, 39)
Treatment effects	Meta-analyses of clinical trial results	(22)
Prevalence of breast density	Prevalence of breast density (BI-RADS a, b, c, d) by age group. Density is assigned at age 40 years and can decrease by one level or remain the same at age 50 years and again at age 65 years.	BCSC
Risk levels for density	Risk of breast cancer based on BI-RADS relative to average density by age group.	BCSC
Comorbidity		
Life tables by comorbidity	Comorbidity-adjusted life expectancy among individuals without cancer	(42)

Note: Not all models use all parameters; some models use parameters as direct inputs and others use them as a target for calibration or other estimation.

Table 2. Digital Mammography Characteristics From the Breast Cancer Surveillance Consortium*

Breast Density	Age	Interval†	Sensitivity–Invasive	Sensitivity–DCIS	Specificity
All	40–49	First	0.846	0.957	0.805
		Annual	0.65	0.928	0.881
		Biennial	0.777	0.949	0.881
	50–64	First	0.909	0.957	0.854
		Annual	0.772	0.928	0.912
		Biennial	0.864	0.949	0.912
	≥65	First	0.94	0.956	0.879
		Annual	0.841	0.926	0.928
		Biennial	0.909	0.948	0.928
BI-RADS a	40–49	First	0.921	0.954	0.872
		Annual	0.806	0.919	0.93
		Biennial	0.881	0.942	0.925
	50–64	First	0.948	0.955	0.903
		Annual	0.868	0.921	0.948
		Biennial	0.921	0.943	0.944
	≥65	First	0.963	0.955	0.916
		Annual	0.903	0.922	0.955
		Biennial	0.943	0.944	0.952
BI-RADS b	40–49	First	0.894	0.948	0.797
		Annual	0.751	0.911	0.884
		Biennial	0.844	0.935	0.876
	50–64	First	0.93	0.949	0.843
		Annual	0.826	0.912	0.912
		Biennial	0.895	0.937	0.906
	≥65	First	0.95	0.95	0.863
		Annual	0.871	0.913	0.924
		Biennial	0.924	0.937	0.919
BI-RADS c	40–49	First	0.817	0.964	0.76
		Annual	0.615	0.937	0.86
		Biennial	0.74	0.955	0.851
	50–64	First	0.876	0.965	0.812
		Annual	0.716	0.938	0.894
		Biennial	0.818	0.956	0.886
	≥65	First	0.909	0.965	0.836
		Annual	0.782	0.938	0.908
		Biennial	0.865	0.956	0.901
BI-RADS d	40–49	First	0.746	0.943	0.815
		Annual	0.512	0.902	0.895
		Biennial	0.652	0.929	0.888
	50–64	First	0.822	0.944	0.857
		Annual	0.623	0.904	0.921
		Biennial	0.747	0.93	0.915
	≥65	First	0.868	0.944	0.876
		Annual	0.702	0.904	0.932
		Biennial	0.808	0.931	0.927

*BCSC, DR285, 2003–2013.
†Screen-detected cancers diagnosed in 1 year; cut-off at next screening.

Table 3. Utility Input Parameter Values for Cancer-Related States

State	Utility	Disutility (Worst Case, 150% and 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

Source: references 45 and 46.

Table 4. Prevalence of Breast Density Group by Age From the Breast Cancer Surveillance Consortium

Age	BI-RADS a Almost Entirely Fat	BI-RADS b Scattered Fibrodens	BI-RADS c Heterogeneously Dense	BI-RADS d Extremely Dense
40–44	0.046	0.338	0.472	0.144
45–49	0.055	0.364	0.458	0.123
50–54	0.075	0.422	0.416	0.086
55–59	0.098	0.471	0.373	0.058
60–64	0.117	0.500	0.338	0.045
65–69	0.130	0.521	0.313	0.036
70–74	0.136	0.537	0.296	0.031
75–79	0.139	0.539	0.291	0.031

Table 5. Relative Risk Levels by BI-RADS Breast Density and Age Group From the Breast Cancer Surveillance Consortium

Age Group	BI-RADS Density	Breast Density–Related Risk*	Risk Levels Associated With Factors Other Than Breast Density†			
			RR, 1	RR, 1.3	RR, 2	RR, 4
40–49	a	0.37	0.37	0.49	0.75	1.49
	b	0.72	0.72	0.93	1.43	2.86
	c	1.16	1.16	1.51	2.32	4.63
	d	1.46	1.46	1.89	2.91	5.83
50–64	a	0.50	0.50	0.65	1.00	2.00
	b	0.84	0.84	1.10	1.69	3.38
	c	1.25	1.25	1.62	2.50	4.99
	d	1.53	1.53	1.99	3.06	6.13
65–74	a	0.61	0.61	0.80	1.22	2.45
	b	0.94	0.94	1.22	1.88	3.75
	c	1.28	1.28	1.66	2.56	5.12
	d	1.45	1.45	1.88	2.90	5.79

*Personal communication with BCSC; referent is average population density.
†Values for each risk/breast density combination are the product of the breast density–related relative risk x the risk level. For example, a group with breast density "d" and an RR of 2 due to a family history of breast cancer has an overall relative risk of breast cancer of 2.91 from ages 40 to 49 years.

BI-RADs density "a" corresponds to almost entirely fatty breasts; "b" to scattered areas of density; "c" to heterogeneously dense; and "d" to extremely dense.

Note: While the models computed results for each of the four density risk groups separately, since there can be variability in radiological interpretation of density level, and breast density legislation in most States groups women into low (BI-RADS a or b) or high (BI-RADS c or d) density, the data below summarizes the weighted density-related risk for breast cancer for low- and high-density groups combined relative to average density.

Age Group	BI-RADS Density	Breast Density–Related Risk	Prevalence
40–49	a and b	0.67	41.2%
	c and d	1.23	58.8%
50–64	a and b	0.79	57.4%
	c and d	1.29	42.7%
65–74	a and b	0.87	68.9%
	c and d	1.30	31.1%

Table 6. Modeling AGE Trial With 13 Years of Followup: Projection of Relative Risk of Breast Cancer Death With Annual Screening at Ages 40 to 49 Years; Biennial Screening at Age 50 and 52 Years vs. Control Biennial Screening at Age 50 and 52 Years

Model	Relative Risk of Breast Cancer Death With 100% Screening
D	0.75
E	0.73
G-E	0.65
M	0.72
S	0.69
W	0.71
Median (range)	0.72 (0.65 to 0.75)

AGE trial invitation (intention to treat) (54): RR, 0.83 (95% CI, 0.66 to 1.04).
AGE trial results for women who were actually screened (54): RR, 0.76 (95% CI, 0.51 to 1.01).

Tables 7a–c. Percent Breast Cancer Mortality Reduction (or Life-Years or Quality-Adjusted Life-Years Gained) and Average Number of Screening Examinations per 1,000 Women by Model and Screening Strategy

Table 7a. Mortality Reduction

Strategies	Average Number of Screenings per 1,000 Women	% Breast Cancer Mortality Reduction per 1,000 Women Screened (vs. No Screening) by Model						Median (Range Across Models)
		D	E	G-E	M	S	W	
B 50–74	11,205	25.6%	26.0%	31.8%	26.8%	24.1%	25.4%	25.8% (24.1–31.8)
B 45–74	13,301	26.6%	27.6%	33.9%	28.4%	25.9%	26.7%	27.2% (25.9–33.9)
H 45–74	16,060	27.7%	29.7%	35.9%	29.2%	27.3%	30.1%	29.5% (27.3–35.9)
B 40–74	16,112	28.3%	30.3%	35.9%	31.9%	28.2%	30.5%	30.4% (28.2–35.9)
H 40–74	20,989	29.0%	32.3%	37.9%	**31.7%**	29.3%	32.8%	32.0% (29.0–37.9)
A 50–74	21,447	32.1%	33.9%	**37.6%**	**27.1%**	**29.1%**	35.3%	33.0% (27.1–37.6)
A 45–74	26,280	34.2%	37.6%	41.6%	**29.4%**	32.3%	39.1%	35.9% (29.4–41.6)
A 40–74	31,194	35.5%	40.1%	43.6%	32.5%	34.4%	42.6%	37.8% (32.5–43.6)

Table 7b. Life-Years Gained

Strategies	Average Number of Screenings per 1,000 Women	Years of Life Gained per 1,000 Women (vs. No Screening) by Model						Median (Range Across Models)
		D	E	G-E	M	S	W	
B 50–74	11,205	153.8	94.0	140.5	146.5	104.2	74.6	122.4 (74.6–153.8)
B 45–74	13,301	168.4	107.7	161.2	171.3	115.2	84.0	138.2 (84.0–171.3)
H 45–74	16,060	175.3	117.9	170.2	171.4	125.1	95.7	147.7 (95.7–175.3)
B 40–74	16,112	183.7	123.7	172.4	194.8	131.6	98.8	152.0 (98.8–194.8)
H 40–74	20,989	191.1	137.6	187.2	211.5	141.0	110.9	164.1 (110.9–211.5)
A 50–74	**21,447**	**180.0**	**125.9**	**167.3**	**156.3**	**133.3**	**104.3**	**144.8 (104.3–180.0)**
A 45–74	26,280	201.3	149.3	196.7	**177.8**	154.2	123.0	166.0 (123.0–201.3)
A 40–74	31,194	217.1	168.8	213.5	218.1	170.1	140.5	191.8 (140.5–218.1)

Table 7c. Quality-Adjusted Life Years Gained

Strategies	Average Number of Screenings per 1,000 Women	Quality-Adjusted Life-Years Gained per 1,000 Women (vs. No Screening) by Model						Median (Range Across Models)
		D	E	G-E	M	S	W	
B 50–74	11,205	114.5	67.3	100.1	109.6	71.9	47.1	86.0 (47.1–114.5)
B 45–74	13,301	123.8	75.6	114.4	129.4	78.8	51.9	96.6 (51.9–129.4)
H 45–74	16,060	126.6	80.9	118.3	128.5	84.5	58.3	101.4 (58.3–128.5)
B 40–74	16,112	133.7	85.4	120.1	148.1	89.1	60.4	104.6 (60.4–148.1)
H 40–74	20,989	134.2	91.0	126.1	159.4	92.5	64.8	109.3 (64.8–159.4)
A 50–74	21,447	**127.0**	**84.1**	**111.4**	**113.2**	**87.5**	**62.4**	**99.5 (62.4–127.0)**
A 45–74	26,280	138.9	97.3	129.5	**129.4**	99.5	71.7	114.5 (71.7–138.9)
A 40–74	31,194	146.6	107.3	137.2	160.6	107.6	80.0	122.4 (80.0–160.6)

A=Annual; B=Biennial; H=Hybrid.

Notes:

1. Average number of mammograms across all models. Not all possible mammograms in the age interval are obtained since some women die from other causes before screening would occur.
2. Model group abbreviations: D (Dana Farber), E (Erasmus Medical Center), G-E (Georgetown University–Einstein University), M (M.D. Anderson Cancer Center), S (Stanford University), W (University of Wisconsin/Harvard).
3. Values in bold show strategies that are dominated ("inefficient") within a specific model; a strategy is classified as dominated if there is another strategy that results in an equal or higher percent mortality reduction/life-years gained/quality-adjusted life-years with fewer average screening examinations.
4. QALYs are adjusted for general health, diagnosis, screening, and treatment.
5. 100% receive recommended age-, stage-, and ER/HER2-specific adjuvant therapy.

Table 8. Benefits of Screening Strategies Based on Starting Ages and Intervals

Scenario	Number of screenings	Potential Benefits (vs. No Screening) Median (Range Across Models)					
		% Breast Cancer Mortality Reduction	Cancer Deaths Averted per 1,000 Women	Life-Years Gained per 1,000 Women	Quality-Adjusted Life-Years Gained per 1,000 Women	Number of All Cancers Diagnosed (Invasive and DCIS) per 1,000 Women	Number of DCIS Cancers Detected per 1,000 Women
Biennial							
50–74	11,127	25.8% (24.1-31.8)	7.1 (3.8-8.7)	122.4 (74.6-153.8)	86.0 (47.1-114.5)	150.7 (134.1-158.5)	26.7 (25.7-32.3)
45–74	13,212	27.2% (25.9-33.9)	7.5 (4.0-9.2)	138.2 (84.0-171.3)	96.6 (51.9-129.4)	150.4 (134.0-158.6)	28.7 (27.7-31.7)
40–74	16,013	30.4% (28.2-35.9)	8.1 (4.5-10.3)	152.0 (98.8-194.8)	104.6 (60.4-148.1)	153.7 (134.2-160.9)	30.5 (30.2-34.7)
Hybrid							
45–74	15,966	29.5% (27.3-35.9)	7.9 (4.5-9.4)	147.7 (95.7-175.3)	101.4 (58.3-128.5)	154.1 (134.2-160.4)	30.2 (29.8-34.8)
40–74	20,884	32.0% (29.0-37.9)	8.5 (4.9-10.3)	164.1 (110.9-211.5)	109.3 (64.8-159.4)	154.9 (134.2-164.4)	32.9 (30.2-36.5)
Annual							
50–74	21,318	33.0% (27.1-37.6)	8.5 (5.3-10.1)	144.8 (104.3-180.0)	99.5 (62.4-127.0)	157.6 (134.5-188.1)	34.3 (30.0-46.3)
45–74	26,136	35.9% (29.4-41.6)	9.3 (5.8-10.7)	166.0 (123.0-201.3)	114.5 (71.7-138.9)	158.7 (134.5-194.0)	38.4 (30.3-48.7)
40–74	31,037	37.8% (32.5-43.6)	10.1 (6.3-11.2)	191.8 (140.5-218.1)	122.4 (80.0-160.6)	159.5 (134.5-197.5)	41.5 (30.3-50.3)

Table 9. Percent of Annual Mortality Reduction Maintained by Biennial Screening by Strategy and Model

Age	Model D	Model E	Model G-E	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%

* Percent of A50–74 maintained by B50–74 = percent mortality reduction B50–74/ percent mortality reduction A50–74.
** Percent of A45–74 maintained by B45–74 = percent mortality reduction B45–74/ percent mortality reduction A45–74.
*** Percent of A40–74 maintained by B40–74 = percent mortality reduction B40–74/ percent mortality reduction A40–74.

Note: Model M does not include a natural history component. It is based on a combination of assumptions regarding underlying incidence trends in the absence of screening that essentially yields a long invasive cancer lead-time; thus, virtually all cancers found with annual screening can also be detected with biennial screening.

Tables 10a and b. Incremental Changes in Benefits by Age of Screening Initiation, Screening Interval, and Model

Table 10a. Start Age of 40 Years (vs. 50 Years)*

	Difference in % Breast Cancer Mortality Reduction		Breast Cancer Deaths Averted		Life-Years Gained per 1,000 Women		QALYs Gained per 1,000 Women	
Model	**Annual**	**Biennial**	**Annual**	**Biennial**	**Annual**	**Biennial**	**Annual**	**Biennial**
D	3.4%	2.7%	1.1	0.9	37.1	29.9	19.6	19.2
E	6.2%	4.3%	1.5	1.0	42.9	29.7	23.2	18.1
G-E	6.0%	4.1%	1.5	1.0	46.2	31.9	25.8	20.0
M	5.3%	5.1%	1.7	1.7	61.8	48.3	47.4	38.5
S	5.2%	4.1%	1.1	0.9	36.8	27.4	20.1	17.2
W	7.3%	5.1%	1.1	0.8	36.2	24.2	17.6	13.3
Median	5.7%	4.2%	1.3	1.0	40.0	29.8	21.7	18.7

* Incremental difference between starting at age 40 or 45 years vs. 50 years.

Table 10b. Start Age of 45 Years (vs. 50 Years)*

	Difference in % Breast Cancer Mortality Reduction		Breast Cancer Deaths Averted		Life-Years Gained per 1,000 Women		QALYs Gained per 1,000 Women	
Model	**Annual**	**Biennial**	**Annual**	**Biennial**	**Annual**	**Biennial**	**Annual**	**Biennial**
D	2.1%	1.0%	0.6	0.3	21.3	14.6	11.9	9.3
E	3.6%	1.6%	0.9	0.4	23.4	13.7	13.2	8.3
G-E	4.0%	2.2%	1.0	0.5	29.4	20.7	18.1	14.3
M	2.3%	1.6%	0.7	0.5	21.5	24.8	16.2	19.8
S	3.1%	1.7%	0.7	0.4	21.5	11.0	12.0	6.9
W	3.8%	1.3%	0.6	0.2	18.7	9.4	9.3	4.8
Median	3.4%	1.6%	0.7	0.4	21.5	14.2	12.6	8.8

* Incremental difference between starting at age 40 or 45 years vs. 50 years.

Annual = comparing A40–74 (or 45–74) to A50–74.
Biennial = comparing B40–74 (or 45–74) to B50–74.
Hybrid = comparing annual screening in the 40s followed by biennial screening from ages 50–74 years to biennial screening from ages 50–74 years; for these incremental comparisons, the hybrid results are the same as the annual results.
100% of women receive adjuvant systemic therapy based on recommended stage- and ER/HER2-specific adjuvant therapy for pre- and post-menopausal women.

Table 11. Harms of Screening Strategies Based on Different Starting Ages and Intervals

Scenario	Number of Screenings	Potential Harms (vs. No Screening) Median (Range Across Models)				
		Number of False-Positives per 1,000 Women	Number of Benign Biopsies per 1,000 Women	Number of Cases of Overdiagnosis (All Cancer) per 1,000 Women	Number of Cases of Overdiagnosis (DCIS) per 1,000 Women	Number of Cases of Overdiagnosis (Invasive) per 1,000 Women
Biennial						
50–74	11,127	953 (830-1325)	145.5 (120.6-205.2)	18.8 (10.8-34.2)	15.6 (9.0-18.8)	2.8 (1.8-15.4)
45–74	13,212	1220 (930-1599)	176.1 (131.1-232.4)	19.1 (11.1-34.0)	15.3 (9.4-20.9)	2.6 (1.6-13.1)
40–74	16,013	1529 (1100-1976)	212.6 (153.2-276.0)	21.4 (12.0-37.7)	17.1 (10.2-23.8)	2.9 (1.8-13.8)
Hybrid						
45–74	15,966	1520 (1160-1968)	202.0 (154.0-266.2)	20.9 (11.6-40.2)	17.1 (9.8-23.5)	2.9 (1.8-16.7)
40–74	20,884	2106 (1480-2623)	256.2 (184.1-325.2)	23.0 (12.1-44.3)	18.1 (10.3-26.6)	2.9 (1.8-17.7)
Annual						
50–74	21,318	1798 (1706-2445)	228.0 (219.3-317.4)	25.1 (11.5-67.9)	22.7 (9.4-32.5)	3.2 (2.1-35.4)
45–74	26,136	2355 (2185-3087)	282.7 (264.6-375.9)	27.9 (12.3-73.8)	24.2 (10.2-37.1)	3.2 (2.1-36.7)
40–74	31,037	2941 (2550-3742)	338.2 (295.8-434.8)	30.0 (12.7-77.3)	25.0 (10.6-39.9)	3.2 (2.1-37.4)

1. Overdiagnosed cases are defined as cases that would not have been clinically detected in the absence of screening; includes DCIS and invasive cancer. Overdiagnosis is calculated by comparing cases detected in the screening scenario to those detected in the unscreened scenario. Model S is excluded since it does not include DCIS. Model M generates very high overdiagnosis based on the assumption that incidence in the absence of screening has essentially remained flat since 1975 to 1979, 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).
2. Number of cancers diagnosed is higher in B50–74 vs. B45–74 because the B50–74 group has a screen at age 74 years, when cancer incidence is the highest. Although the B45–74 group has two more screens than B50–74, the last screen for B45–74 is at age 73 years.
3. 100% of women receive adjuvant systemic therapy based on recommended stage- and ER/HER2-specific adjuvant therapy for pre- and post-menopausal women.

Table 12. Percent of Cases (Invasive Cancer and DCIS) Overdiagnosed by Strategy

Scenario	Percent of Cases That Are Overdiagnosis per 1,000 Women Screened Median (Range Across Models)*
Biennial	
50–74	11.9% (7.9–22.2)
45–74	12.0% (8.1–22.0)
40–74	13.3% (8.7–23.9)
Hybrid	
45–74	13.0% (8.4–25.1)
40–74	14.2% (8.7–26.9)
Annual	
50–74	15.3% (8.3–36.1)
45–74	16.7% (8.9–38.0)
40–74	17.7% (9.1–39.1)

Percent of all cases diagnosed under the screening strategy that are overdiagnosed.

*The upper range is based on results from one model that essentially assumes incidence in the absence of screening remained fairly flat since 1975 to 1979; hence all increases are 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 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.

Table 13a. Benefits and Harms by Breast Density, Risk Level, and Screening Strategy

Screening Strategy	Number of Mammograms*	Percent Mortality Reduction Median (Range)	Breast Cancer Deaths Averted per 1,000 Women Screened Median (Range)	Life-Years Saved per 1,000 Women Screened Median (Range)	False-Positives per 1,000 Women Screened Median (Range)	Number of Benign Biopsies per 1,000 Women Screened Median (Range)	Cases Overdiagnosed per 1,000 Women Screened Median (Range)
B50–74							
RR, 1	11,095	25.4 (22.7-31.8)	5.5 (3.8-7.8)	82.0 (74.6-140.2)	1018 (1015-1325)	151.0 (150.4-205.2)	17.6 (10.8-25.5)
B40–74							
RR, 1.3	15,742	30.9 (28.6-36.1)	8.5 (5.9-11.3)	151.1 (127.3-220.7)	1723 (1627-1951)	240.4 (225.3-272.6)	25.7 (12.0-35.1)
RR, 2	15,261	31.2 (29.5-36.5)	12.1 (8.7-16.6)	217.7 (189.4-327.8)	1677 (1566-1895)	234.0 (216.8-264.7)	35.0 (11.9-48.7)
RR, 4	14,146	33.5 (31.4-37.4)	19.6 (16.3-29.4)	365.3 (354.5-591.9)	1571 (1409-1748)	219.1 (195.0-244.2)	50.5 (12.3-73.3)
Hybrid 45–74							
RR, 1.3	15,695	30.4 (27.6-36.0)	8.2 (5.8-11.3)	139.9 (122.6-217.5)	1595 (1574-1942)	210.2 (207.5-262.6)	25.1 (11.6-35.0)
RR, 2	15,202	30.8 (28.3-36.5)	11.6 (8.6-16.6)	200.7 (182.8-323.2)	1548 (1511-1882)	203.9 (198.8-254.4)	34.0 (11.6-48.6)
RR, 4	14,060	32.9 (29.8-37.3)	18.6 (16.0-29.3)	340.0 (330.6-581.6)	1439 (1350-1725)	189.1 (177.0-233.0)	48.4 (12.1-73.1)
Hybrid 40–74							
RR, 1.3	20,570	33.6 (30.9-38.1)	9.2 (6.4-11.9)	170.7 (144.2-239.4)	2289 (2175-2595)	281.4 (263.9-321.5)	27.8 (12.1-37.1)
RR, 2	20,021	34.3 (32.0-38.6)	13.1 (9.5-17.6)	247.2 (216.0-356.2)	2235 (2103-2531)	274.3 (254.5-313.1)	38.1 (12.0-51.6)
RR, 4	18,733	37.0 (34.4-39.7)	21.4 (18.0-31.2)	419.3 (408.3-646.5)	2109 (1921-2363)	257.7 (230.8-291.0)	55.9 (12.4-77.7)
A 50-74							
RR, 1.3	20,827	35.5 (30.6-37.7)	9.2 (6.7-11.8)	141.3 (133.0-212.9)	1780 (1767-2402)	223.1 (221.4-312.1)	29.9 (11.7-43.3)
RR, 2	19,960	36.2 (31.0-37.9)	12.7 (10.1-17.3)	199.3 (198.5-314.6)	1706 (1669-2305)	214.2 (209.5-300.0)	39.9 (12.0-59.9)
RR, 4	18,014	37.8 (31.2-38.3)	19.4 (18.4-30.2)	362.7 (309.8-557.7)	1541 (1427-2055)	194.1 (180.0-268.7)	54.3 (13.2-89.5)
A 40–74							
RR, 1.3	30,417	43.0 (38.5-43.9)	11.5 (8.1-13.7)	208.5 (180.2-273.2)	3061 (2952-3694)	354.2 (337.6-429.1)	36.3 (12.8-48.3)
RR, 2	29,379	43.9 (39.6-44.4)	16.3 (12.1-20.2)	299.7 (270.3-406.0)	2969 (2833-3585)	343.3 (323.6-416.3)	49.3 (13.0-66.8)
RR, 4	27,003	45.6 (41.8-46.5)	26.0 (22.7-35.8)	501.9 (498.9-735.3)	2756 (2534-3299)	318.3 (288.5-382.7)	70.9 (14.0-99.6)

*Number of mammograms is the median across models; as risk increases, more women develop and die of breast cancer, therefore the number of lifetime screening mammograms decreases.

Table 13b. Ratio of Harms to Benefits by Risk Level (Median and Range Across Three Models–E, G-E, W)

Screening Strategy	Number of Mammograms*/Death Averted per 1,000 Women Screened	False-Positives/Death Averted per 1,000 Women Screened	Number of Benign Biopsies/Death Averted per 1,000 Women Screened	Overdiagnosed Cases/Death Averted per 1,000 Women Screened
	Median (Range)	Median (Range)	Median (Range)	Median (Range)
B50–74				
RR, 1	2024 (1428-2920)	185.8 (169.5-268.0)	27.6 (26.2-39.7)	3.2 (1.4-6.7)
B40-74				
RR, 1.3	1843 (1407-2681)	201.8 (172.9-278.0)	28.1 (24.2-38.5)	3.0 (1.1-6.0)
RR, 2	1259 (929-1747)	138.4 (114.0-180.3)	19.3 (15.9-25.0)	2.9 (0.7-5.6)
RR, 4	723 (486-848)	80.3 (59.4-86.4)	11.2 (8.3-12.0)	2.6 (0.4-4.5)
Hybrid 45-74				
RR, 1.3	1908 (1404-2715)	193.9 (172.4-273.3)	25.6 (23.3-36.0)	3.1 (1.0-6.1)
RR, 2	1305 (925-1766)	132.9 (113.3-176.7)	17.5 (15.3-23.3)	2.9 (0.7-5.7)
RR, 4	757 (482-856)	77.5 (58.8-84.3)	10.2 (7.9-11.1)	2.6 (0.4-4.6)
Hybrid 40–74				
RR, 1.3	2233 (1740-3226)	248.5 (217.8-341.7)	30.5 (27.0-41.5)	3.0 (1.0-5.8)
RR, 2	1524 (1152-2089)	170.2 (144.1-220.5)	20.9 (17.8-26.7)	2.9 (0.7-5.4)
RR, 4	875 (608-1022)	98.5 (75.8-106.7)	12.0 (9.3-12.8)	2.6 (0.4-4.3)
A 50-74				
RR, 1.3	2276 (1789-3073)	204.0 (194.5-262.5)	26.5 (24.4-32.9)	3.3 (1.0-6.4)
RR, 2	1568 (1172-1957)	134.0 (133.4-165.9)	17.4 (16.8-20.8)	3.1 (0.7-6.0)
RR, 4	927 (601-933)	77.5 (68.1-79.3)	9.8 (8.9-10.0)	2.8 (0.4-4.9)
A 40–74				
RR, 1.3	2646 (2246-3726)	269.5 (266.3-362.8)	31.3 (30.8-41.5)	3.2 (0.9-5.9)
RR, 2	1806 (1479-2388)	182.5 (177.4-231.9)	21.1 (20.6-26.5)	3.0 (0.6-5.5)
RR, 4	1038 (768-1158)	105.9 (92.1-111.7)	12.2 (10.7-12.7)	2.7 (0.4-4.4)

*Number of mammograms is the median across models.

Table 13c. Benefits and Harms by Breast Density Group (Median and Range Across Three Models–E, G-E, W)

Screening Strategy	Number of Mammograms*	Percent Mortality Reduction	Breast Cancer Deaths Averted per 1,000 Women Screened	Life Years Saved per 1,000 Women Screened	False-Positives per 1,000 Women Screened	Benign Biopsies per 1,000 Women Screened	Cases Overdiagnosed per 1,000 Women Screened
		Median (Range)	Median (Range)	Median (Range)	Median (Range)	Median (Range)	Median (Range)
B50–74							
Low density†	11246	31.5 (25.3-39.2)	3.9 (3.0-5.7)	42.9 (42.8-72.5)	923 (869-1154)	136.8 (128.9-179.2)	14.6 (11.0-23.5)
High density	11083	23.2 (22.4-31.2)	6.3 (3.9-8.5)	95.7 (77.7-152.1)	1198 (1138-1425)	177.7 (168.7-220.7)	20.3 (10.7-23.4)
B40–74							
Low density	16148	37.4 (32.4-44.6)	5.0 (3.5-6.5)	68.1 (58.1-92.9)	1551 (1500-1742)	195.8 (190.3-223.2)	17.3 (12.3-25.9)
High density	15973	27.9 (27.7-35.4)	7.9 (4.8-9.6)	139.7 (104.1-188.3)	1998 (1817-2115)	289.1 (260.0-305.7)	24.0 (11.9-26.2)
Hybrid 45-74							
Low density	16109	36.9 (30.9-44.2)	4.8 (3.5-6.5)	61.7 (56.0-89.8)	1453 (1429-1737)	184.1 (179.9-224.0)	16.9 (11.8-25.7)
High density	15915	27.4 (26.8-35.4)	7.6 (4.7-9.6)	129.7 (99.7-186.2)	1860 (1761-2106)	255.2 (241.6-297.4)	23.4 (11.5-26.2)
Hybrid 40–74							
Low density	21037	40.0 (35.1-46.7)	5.4 (3.8-6.8)	77.9 (65.4-100.1)	2083 (1996-2326)	238.7 (229.7-271.0)	18.6 (12.3-27.2)
High density	20839	30.7 (29.9-37.5)	8.5 (5.2-10.2)	158.0 (119.3-205.4)	2629 (2403-2802)	341.2 (307.2-364.8)	25.8 (12.0-28.1)
A50-74							
Low density	21540	42.3 (34.3-45.3)	5.3 (4.0-6.6)	59.5 (58.1-84.3)	1654 (1554-2124)	191.4 (179.9-257.2)	20.3 (11.5-30.3)
High density	21255	32.7 (30.1-37.0)	8.5 (5.6-10.1)	132.2 (110.6-181.9)	2133 (2023-2630)	267.2 (253.4-341.5)	28.4 (11.5-34.1)
A40-74							
Low density	31292	51.0 (43.9-52.8)	6.8 (4.8-7.7)	94.3 (81.0-112.0)	2764 (2730-3294)	287.7 (283.0-345.5)	24.4 (12.8-33.9)
High density	30999	40.1 (37.5-43.3)	10.6 (6.8-11.8)	193.8 (152.0-235.0)	3556 (3291-4005)	428.8 (391.3-482.7)	33.9 (12.7-38.4)

†Low density= BI-RADS a and b; high density= BI-RADS c and d

*Number of mammograms is the median across models; as density increases, more women develop and die of breast cancer, therefore the number of lifetime screening mammograms decreases.

Table 13d. Ratio of Harms to Benefits by Breast Density Group† (Median and Range Across Three Models–E, G-E, W)

Screening Strategy	Number of Mammograms*/ Death Averted per 1,000 Women Screened	False-Positives/Death Averted per 1,000 Women Screened	Benign Biopsies/Death Averted per 1,000 Women Screened	Overdiagnosed Cases/Death Averted per 1,000 Women Screened
	Median (Range)	Median (Range)	Median (Range)	Median (Range)
B50–74				
Low density	2709 (1975-3777)	223.4 (201.0-309.8)	33.1 (31.2-45.9)	3.7 (1.9-7.9)
High density	1890 (1309-2779)	188.7 (168.3-288.1)	28.0 (26.1-42.7)	3.2 (1.3-5.9)
B40–74				
Low density	3092 (2486-4575)	310.8 (266.5-425.0)	39.2 (34.1-53.9)	3.5 (1.9-7.3)
High density	2147 (1661-3336)	254.5 (219.9-382.6)	36.8 (31.8-54.7)	3.1 (1.2-5.5)
Hybrid 45-74				
Low density	3231 (2504-4625)	300.0 (268.2-417.1)	37.8 (34.6-52.8)	3.6 (1.8-7.4)
High density	2214 (1653-3384)	244.9 (218.8-377.4)	33.6 (30.9-51.8)	3.1 (1.2-5.6)
Hybrid 40–74				
Low density	3764 (3092-5568)	385.9 (339.9-528.2)	44.2 (39.6-60.8)	3.5 (1.8-7.2)
High density	2565 (2046-3968)	310.7 (275.1-460.9)	40.3 (35.8-58.9)	3.0 (1.2-5.4)
A50-74				
Low density	3823 (3280-5390)	319.9 (294.3-413.9)	38.7 (34.1-47.9)	3.8 (1.7-7.6)
High density	2691 (2111-3774)	261.2 (249.9-363.9)	33.9 (31.3-45.6)	3.3 (1.1-6.1)
A40-74				
Low density	4422 (4080-6499)	425.6 (408.6-566.9)	44.6 (42.5-58.8)	3.6 (1.7-7.0)
High density	3073 (2633-4496)	340.2 (335.3-482.9)	41.0 (40.4-57.4)	3.2 (1.1-5.6)

†Low density= BI-RADS a and b; high density= BI-RADS c and d
*Number of mammograms is the median across models.

Table 13e1. Benefits by Risk Level and Breast Density Group† (Median and Range Across Three Models–E, G-E, W)

Screening Strategy	Number of Mammograms*		Percent Mortality Reduction		Breast Cancer Deaths Averted per 1,000 Women Screened		Life-Years Saved per 1,000 Women Screened	
	Median		Median (Range)		Median (Range)		Median (Range)	
	Low Density	High Density	Low Density	High Density	Low Density	High Density	Low Density	High Density
B50–74								
RR, 1	11246	11083	31.5 (25.3-39.2)	23.2 (22.4-31.2)	3.9 (3.0-5.7)	6.3 (3.9-8.5)	42.9 (42.8-72.5)	95.7 (77.7-152.1)
B40–74								
RR, 1.3	15973	15755	37.6 (32.6-44.7)	28.2 (28.1-35.6)	6.3 (4.6-8.4)	9.8 (6.1-12.2)	87.1 (75.1-119.8)	176.0 (132.1-240.4)
RR, 2	15574	15261	37.8 (33.1-44.7)	29.1 (28.7-35.9)	9.1 (6.8-12.6)	13.9 (9.0-17.9)	127.7 (113.1-180.2)	252.7 (197.7-355.6)
RR, 4	14520	13971	38.7 (34.1-44.8)	31.0 (30.6-36.8)	15.5 (12.9-23.2)	22.0 (16.8-31.4)	225.0 (215.0-339.7)	419.1 (368.7-636.7)
Hybrid 45-74								
RR, 1.3	15928	15685	37.0 (31.1-44.3)	27.7 (27.2-35.6)	6.0 (4.5-8.3)	9.5 (6.0-12.2)	78.8 (71.8-115.9)	163.2 (127.4-237.5)
RR, 2	15518	15164	37.0 (31.5-44.2)	28.2 (28.0-35.9)	8.7 (6.7-12.4)	13.3 (8.9-17.9)	115.1 (107.7-174.2)	233.2 (190.2-350.9)
RR, 4	14430	13800	37.8 (32.1-44.3)	30.1 (29.5-36.7)	14.6 (12.6-22.9)	20.9 (16.4-31.4)	203.5 (199.7-327.4)	378.9 (352.0-626.3)
Hybrid 40-74								
RR, 1.3	20840	20595	40.0 (35.3-46.8)	31.0 (30.4-37.7)	6.9 (4.8-8.8)	10.6 (6.7-13.0)	99.8 (83.9-129.4)	199.3 (152.5-262.6)
RR, 2	20390	20041	40.7 (35.9-46.8)	31.8 (31.5-38.2)	9.9 (7.4-13.2)	15.0 (10.0-19.1)	146.7 (127.7-194.9)	287.9 (228.3-388.9)
RR, 4	19196	18584	41.9 (37.3-47.0)	34.3 (34.0-39.3)	16.9 (14.0-24.4)	24.1 (18.7-33.5)	260.6 (244.1-368.2)	483.0 (429.9-701.3)
A 50-74								
RR, 1.3	21205	20847	42.4 (34.3-45.4)	33.0 (30.4-37.2)	6.7 (5.1-8.5)	10.6 (7.1-12.8)	75.4 (74.5-108.7)	164.7 (140.9-232.0)
RR, 2	20450	19929	42.4 (34.2-45.2)	33.6 (30.7-37.3)	9.5 (7.7-12.7)	14.6 (10.6-18.6)	111.9 (108.3-162.9)	230.0 (209.0-340.3)
RR, 4	18474	17562	42.9 (33.5-44.9)	34.8 (30.8-37.7)	15.2 (14.3-23.2)	21.9 (19.1-32.2)	209.8 (179.1-303.0)	375.6 (352.9-597.9)
A 40–74								
RR, 1.3	30906	30549	51.0 (44.2-52.8)	40.4 (38.0-43.6)	8.6 (6.2-9.9)	13.2 (8.7-15.0)	120.4 (103.8-144.3)	243.7 (193.4-300.5)
RR, 2	30031	29533	51.2 (44.6-52.9)	41.4 (39.1-44.1)	12.3 (9.3-14.9)	18.6 (13.0-22.0)	176.1 (156.4-217.5)	349.2 (289.6-444.7)
RR, 4	27738	26886	52.2 (45.6-53.1)	43.8 (41.3-45.4)	20.7 (17.4-27.5)	29.3 (24.0-38.7)	308.2 (296.2-410.4)	574.0 (535.8-799.9)

†Low density= BI-RADS a and b; high density= BI-RADS c and d

*Number of mammograms is the median across models; as risk (and density) increases, more women develop and die of breast cancer, therefore the number of lifetime screening mammograms decreases.

Table 13e2. Harms by Risk Level and Breast Density Group† (Median and Range Across Three Models–E, G-E, W)

Screening Strategy	Number of Mammograms*		False-Positives per 1,000 Women Screened		Benign Biopsies per 1,000 Women Screened		Cases Overdiagnosed per 1,000 Women Screened	
	Median		Median (Range)		Median (Range)		Median (Range)	
	Low Density	High Density	Low Density	High Density	Low Density	High Density	Low Density	High Density
B50–74								
RR, 1	11246	11083	923 (869-1154)	1198 (1138-1425)	136.8 (128.9-179.2)	177.7 (168.7-220.7)	14.6 (11.0-23.5)	20.3 (10.7-23.4)
B40–74								
RR, 1.3	15973	15755	1536 (1482-1727)	1972 (1784-2085)	193.7 (187.8-221.1)	285.4 (255.2-301.5)	21.5 (12.2-32.6)	29.4 (11.8-32.4)
RR, 2	15574	15261	1503 (1441-1692)	1915 (1708-2018)	189.3 (182.3-216.3)	277.4 (244.6-292.0)	29.9 (12.2-46.5)	39.6 (11.8-44.3)
RR, 4	14520	13971	1424 (1333-1597)	1783 (1518-1843)	178.5 (167.7-203.5)	258.9 (217.9-267.3)	45.5 (12.4-74.8)	55.7 (12.1-65.0)
Hybrid 45-74								
RR, 1.3	15928	15685	1434 (1414-1720)	1833 (1726-2074)	181.6 (177.8-221.8)	251.5 (236.8-292.9)	21.0 (11.8-32.5)	28.6 (11.5-32.4)
RR, 2	15518	15164	1392 (1380-1683)	1774 (1648-2003)	175.9 (173.3-216.8)	243.4 (226.1-283.0)	29.1 (11.8-46.3)	38.4 (11.5-44.4)
RR, 4	14430	13800	1298 (1281-1582)	1640 (1453-1816)	162.4 (161.1-203.3)	225.0 (199.1-256.9)	44.1 (12.2-74.5)	53.3 (12.0-65.1)
Hybrid 40-74								
RR, 1.3	20840	20595	2065 (1975-2308)	2598 (2364-2768)	236.4 (227.0-268.8)	337.1 (301.9-360.2)	23.2 (12.3-34.3)	31.7 (11.9-34.7)
RR, 2	20390	20041	2026 (1927-2269)	2532 (2277-2691)	231.4 (220.8-263.6)	328.2 (290.2-349.9)	32.5 (12.2-48.9)	43.0 (11.9-47.6)
RR, 4	19196	18584	1932 (1803-2161)	2376 (2055-2490)	219.2 (204.8-249.7)	307.3 (260.8-323.0)	50.2 (12.5-78.8)	61.4 (12.3-69.8)
A 50-74								
RR, 1.3	21205	20847	1625 (1529-2098)	2089 (1969-2579)	188.2 (177.2-254.1)	262.0 (246.9-335.2)	25.1 (11.6-38.2)	34.5 (11.6-42.1)
RR, 2	20450	19929	1559 (1476-2037)	1997 (1851-2464)	180.8 (171.3-247.1)	250.8 (232.4-320.8)	34.3 (11.9-54.4)	45.6 (12.1-57.6)
RR, 4	18474	17562	1390 (1351-1875)	1794 (1558-2168)	161.8 (157.3-228.5)	225.9 (196.6-283.7)	49.9 (13.0-87.1)	60.2 (13.3-83.7)
A 40–74								
RR, 1.3	30906	30549	2733 (2694-3265)	3503 (3226-3948)	284.3 (279.2-342.3)	422.4 (383.5-475.8)	30.3 (12.8-42.6)	41.4 (12.7-47.5)
RR, 2	30031	29533	2666 (2614-3197)	3388 (3080-3817)	277.2 (270.6-335.0)	408.7 (366.1-460.3)	42.1 (13.0-60.6)	55.8 (13.0-64.9)
RR, 4	27738	26886	2505 (2406-3013)	3127 (2717-3478)	259.9 (248.2-315.2)	377.6 (322.9-419.8)	64.0 (13.8-96.9)	78.0 (14.0-94.1)

†Low density= BI-RADS a and b; high density= BI-RADS c and d

*Number of mammograms is the median across models; as risk (and density) increases, more women develop and die of breast cancer, therefore the number of lifetime screening mammograms decreases.

Table 13f. Ratio of Harms to Benefits by Risk Level and Breast Density Group† (Median and Range Across Three Models–E, G-E, W)

Screening Strategy	Number of Mammograms*/Death Averted per 1,000 Women Screened		False-Positives/Death Averted per 1,000 Women Screened		Benign Biopsies/Death Averted per 1,000 Women Screened		Overdiagnosed Cases/Death Averted per 1,000 Women Screened	
	Median (Range)		Median (Range)		Median (Range)		Median (Range)	
	Low Density	High Density	Low Density	High Density	Low Density	High Density	Low Density	High Density
B50-74								
RR, 1	2709 (1975-3777)	1890 (1309-2779)	223.4 (201.0-309.8)	188.7 (168.3-288.1)	33.1 (31.2-45.9)	28.0 (26.1-42.7)	3.7 (1.9-7.9)	3.2 (1.3-5.9)
B40-74								
RR, 1.3	2412 (1917-3509)	1691 (1287-2571)	242.9 (205.4-325.5)	201.0 (170.4-294.2)	30.6 (26.3-41.3)	29.1 (24.6-42.1)	3.4 (1.5-7.2)	3.0 (1.0-5.3)
RR, 2	1626 (1256-2274)	1157 (851-1663)	164.4 (134.6-210.4)	138.2 (112.5-189.3)	20.7 (17.2-26.6)	20.0 (16.3-27.1)	3.3 (1.0-6.8)	2.9 (0.7-4.9)
RR, 4	902 (644-1124)	670 (445-807)	92.2 (68.9-103.2)	81.0 (58.7-90.5)	11.6 (8.8-13.0)	11.8 (8.5-13.0)	2.9 (0.5-5.8)	2.5 (0.4-3.9)
Hybrid 45-74								
RR, 1.3	2520 (1929-3559)	1745 (1281-2600)	234.2 (206.5-320.4)	193.3 (169.4-289.1)	29.5 (26.6-40.6)	26.5 (23.9-39.7)	3.5 (1.4-7.3)	3.0 (0.9-5.4)
RR, 2	1701 (1265-2314)	1197 (845-1682)	158.5 (135.2-207.6)	133.0 (111.6-185.8)	19.9 (17.4-26.2)	18.2 (15.8-25.5)	3.3 (0.9-6.9)	2.9 (0.6-5.0)
RR, 4	951 (647-1144)	701 (440-814)	89.1 (69.0-101.6)	78.4 (57.9-88.3)	11.1 (8.9-12.8)	10.8 (8.2-12.1)	3.0 (0.5-5.9)	2.5 (0.4-4.0)
Hybrid 40-74								
RR, 1.3	2934 (2386-4298)	2023 (1587-3053)	301.2 (262.2-407.3)	245.6 (213.3-353.9)	34.5 (30.5-46.8)	31.9 (27.8-45.2)	3.4 (1.4-7.1)	3.0 (0.9-5.2)
RR, 2	1980 (1568-2767)	1383 (1052-1976)	203.9 (172.2-261.6)	168.6 (141.3-228.0)	23.3 (20.0-30.0)	21.8 (18.4-29.1)	3.3 (0.9-6.6)	2.9 (0.6-4.8)
RR, 4	1098 (809-1370)	800 (554-963)	114.1 (88.7-128.7)	98.6 (74.3-109.7)	12.9 (10.3-14.6)	12.8 (9.6-13.9)	3.0 (0.5-5.6)	2.5 (0.4-3.7)
A50-74								
RR, 1.3	2985 (2520-4135)	2125 (1629-2883)	245.7 (229.8-316.8)	201.6 (197.4-277.1)	29.8 (26.6-36.7)	26.2 (24.8-34.7)	3.8 (1.4-7.4)	3.3 (0.9-5.9)
RR, 2	2026 (1644-2662)	1469 (1069-1839)	160.1 (156.1-203.0)	136.6 (132.1-175.4)	19.4 (18.1-23.5)	17.2 (17.2-22.0)	3.6 (0.9-7.1)	3.1 (0.6-5.5)
RR, 4	1153 (831-1288)	876 (545-879)	88.9 (80.7-96.9)	81.7 (67.3-82.0)	10.4 (9.8-11.3)	10.3 (8.8-10.3)	3.3 (0.6-6.1)	2.7 (0.4-4.4)
A40-74								
RR, 1.3	3444 (3146-5000)	2419 (2036-3453)	328.2 (318.8-435.9)	264.6 (263.1-370.4)	34.4 (33.2-45.2)	31.9 (31.7-44.0)	3.5 (1.3-6.9)	3.1 (0.8-5.5)
RR, 2	2323 (2058-3234)	1654 (1341-2216)	215.9 (214.7-281.5)	181.8 (173.4-236.9)	22.5 (22.4-29.1)	21.9 (20.9-28.2)	3.4 (0.9-6.5)	3.0 (0.6-5.0)
RR, 4	1289 (1050-1590)	958 (694-1070)	121.0 (109.6-137.9)	106.6 (89.8-113.3)	12.6 (11.5-14.2)	12.9 (10.8-13.5)	3.1 (0.5-5.6)	2.7 (0.4-3.9)

†Low density = BI-RADS a and b; high density = BI-RADS c and d

*Number of mammograms is the median across models.

Table 14. Example of Comorbidity Prevalence and Remaining Life Expectancy at Age 74 Years

Level	Types of Diseases	Prevalence	Remaining Life Expectancy (Years)
None	None	69%	17
Mild	History of myocardial infarction, acute myocardial infarction, ulcer, or rheumatologic disease	2%	15
Moderate	(Cardio)vascular disease; paralysis; diabetes	12%	13
Severe	AIDS; mild or severe liver disease; chronic renal failure	17%	9

Source: reference 42.

Table 15. Impact of Disutilities on Screening Benefits by Screening Strategy

	Undiscounted Benefits per 1,000 Women Screened vs. No Screening Model Median (Range Across Models)							
	B50–74	**B45–74**	**B40–74**	**H45–74**	**H40–74**	**A50–74**	**A45–74**	**A40–74**
LYS	122.4 (74.6–153.8)	138.2 (84.0–171.3)	152.0 (98.8–194.8)	147.7 (95.7–175.3)	164.1 (110.9–211.5)	144.8 (104.3–180.0)	166.0 (123.0–201.3)	191.8 (140.5–218.1)
QALYs with disutility of age-specific general health	91.9 (56.9–116.9)	104.4 (64.5–130.4)	115.0 (76.1–149.2)	111.5 (73.5–133.8)	124.4 (85.7–161.8)	108.8 (79.6–136.4)	125.3 (94.4–153.2)	146.3 (108.2–167.2)
Percent change from LYS	24.2 (23.7–25.7)	23.8 (23.3–25.3)	23.5 (23.0–25.2)	23.7 (23.2–25.3)	23.4 (22.7–25.0)	24.2 (23.7–25.7)	23.8 (23.3–25.4)	23.5 (22.9–25.1)
QALYs with general health + screening	90.8 (55.9–115.8)	103.1 (63.2–129.1)	113.5 (74.5–147.7)	110.0 (72.0–132.3)	122.4 (83.7–159.8)	106.7 (77.6–134.4)	122.8 (91.9–150.7)	143.3 (105.3–164.2)
Percent change from prior QALY	1.2 (0.9–1.8)	1.3 (1.0–1.9)	1.4 (1.0–2.0)	1.4 (1.2–2.1)	1.7 (1.2–2.3)	1.9 (1.5–2.5)	2.0 (1.6–2.6)	2.1 (1.7–2.7)
QALYs with general health + screening + diagnosis	80.8 (46.9–108.8)	90.8 (51.5–128.3)	98.4 (59.9–146.8)	95.3 (57.8–127.9)	102.7 (64.2–158.8)	93.0 (61.8–118.9)	107.1 (70.9–131.8)	114.4 (79.0–162.9)
Percent change from prior QALY	11.0 (0.7–16.0)	11.9 (0.6–18.5)	13.2 (0.6–19.6)	13.3 (0.7–19.7)	16.1 (0.6–23.3)	15.7 (1.0–20.3)	17.4 (1.0–22.8)	19.4 (0.8–25.0)
QALYs with general health + screening + diagnosis + treatment	86.0 (47.1–114.5)	96.6 (51.9–129.4)	104.6 (60.4–148.1)	101.4 (58.3–128.5)	109.3 (64.8–159.4)	99.5 (62.4–127.0)	114.5 (71.7–138.9)	122.4 (80.0–160.6)
Percent change from prior QALY	-6.2 (-8.6 to -0.5)	-6.2 (-7.8 to -0.8)	-6.3 (-7.1 to -0.8)	-6.3 (-7.5 to -0.5)	-6.4 (-7.0 to -0.4)	-6.6 (-7.4 to 1.9)	-6.5 (-7.3 to 1.8)	-6.3 (-7.3 to 1.4)

Percent change from row above.