Published in final edited form as:
Ann Intern Med. 2009 November 17; 151(10): 738–747. doi: 10.1059/0003-4819-151-10-200911170-00010.

EFFECTS OF MAMMOGRAPHY SCREENING UNDER DIFFERENT SCREENING SCHEDULES: MODEL ESTIMATES OF POTENTIAL BENEFITS AND HARMS

Jeanne S. Mandelblatt, MD, MPH,
Departments of Oncology and Medicine, Georgetown University Medical Center and Cancer Control Program, Lombardi Comprehensive Cancer Center, Washington, DC, USA

Kathleen A. Cronin, Ph.D.,
Surveillance Research Program , Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Stephanie Bailey, Ph.D.,
Department of Radiology, School of Medicine, Stanford University, Palo Alto, California, USA

Donald A. Berry, Ph.D.,
Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Harry J. de Koning, MD, Ph.D.,
Erasmus MC, Department of Public Health, Rotterdam, the Netherlands

Gerrit Draisma, Ph.D.,
Erasmus MC, Department of Public Health, Rotterdam, the Netherlands

Hui Huang, MS,
Dana-Farber Cancer Institute, Boston, Mass, USA

Sandra J. Lee, D.Sc.,
Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Mass, USA

Mark Munsell, MS,
Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Sylvia K. Plevritis, Ph.D.,
Department of Radiology, School of Medicine, Stanford University, Palo Alto, California, USA

Peter Ravdin, MD, Ph.D.,
Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Clyde B. Schechter, MD, MA,
Departments of Family & Social Medicine and Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY, USA

Please address correspondence to: Jeanne Mandelblatt, MD, MPH, Lombardi Comprehensive Cancer Center, 3300 Whitehaven Street, NW, suite 4100, Washington, DC 20007, Telephone 202-687-0812/ facsimile 202-687-0305, mandelbj@georgetown.edu.

*This work was done by six independent modeling teams from Dana-Farber Cancer Institute (PI: Lee), Erasmus University (PI: de Koning), Georgetown University Medical Center, Lombardi Comprehensive Cancer Center (PI: Mandelblatt), Harvard Medical School, Harvard Pilgrim Health Care/University of Wisconsin (PI: Stout), MD Anderson Comprehensive Cancer Center (PI: Berry) and Stanford University (PI: Plevritis). Jeanne Mandelblatt and Kathy Cronin were the writing and coordinating committee for the project; all other collaborators are listed in alphabetical order; Eric Feuer was responsible for overall CISNET project direction.

Bronislava Sigal, Ph.D.,
Department of Radiology, School of Medicine, Stanford University, Palo Alto, California, USA

Michael A. Stoto, Ph.D.,
Department of Health Systems Administration, School of Nursing and Health Studies, Georgetown University Medical Center, Washington, DC

Natasha K. Stout, Ph.D.,
Department of Ambulatory Care and Prevention at Harvard Medical School and Harvard Pilgrim Health Care, Boston, Mass, USA

Nicolien T van Ravesteyn, MSc,
Erasmus MC, Department of Public Health, Rotterdam, the Netherlands

John Venier, MS,
Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Marvin Zelen, Ph.D.,
Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Mass, USA

Eric J. Feuer, Ph.D., and
Surveillance Research Program , Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

The Breast Cancer Working Group of the Cancer Intervention and Surveillance Modeling Network (CISNET)*

Abstract

Background—Despite trials of mammography and widespread use, optimal screening policy is controversial.

Design and Objective—Six models use common data elements to evaluate US screening strategies.

Data Sources—The models use national data on age-specific incidence, competing mortality, mammography characteristics and treatment effects.

Target Population and Time Horizon—A contemporary population cohort followed over their lifetimes.

Perspective—We use a societal perspective for analysis.

Interventions—We evaluate 20 screening strategies with varying initiation and cessation ages applied annually or biennially.

Outcome Measures—Number of mammograms, breast cancer mortality reduction or life years gained [LYG] (vs. no screening), false positives, unnecessary biopsies and over-diagnosis.

Results of Base Case—The 6 models produce consistent rankings of screening strategies. Screening biennially maintains an average of 81% (range across strategies and models 67–99%) of the benefit of annual screening with almost half the number of false positives. Screening biennially from ages 50 to 69 achieves a median 16.5% (range 15%–23%) breast cancer mortality reduction vs. no screening. Initiating biennial screening at age 40 (vs. 50) reduces mortality by an additional 3% (range 1%–6%), consumes more resources and yields more false positives. Biennial screening after age 69 yields some additional mortality reduction in all models but over-diagnosis increases most substantially at older ages.

Sensitivity Analysis Results—Varying test sensitivity or treatment patterns do not change conclusions.

Limitations—Results do not include morbidity from false positives, knowledge of earlier diagnosis or under-going unnecessary treatment.

Conclusion—Biennial screening achieves most of the benefit of annual screening with less harm. Decisions about the best strategy depend on program and individual objectives and the weight placed on benefits, harms and resource considerations.

INTRODUCTION

In 2009 an estimated 193,370 women in the United States (US) will develop invasive breast cancer and about 40,170 of them will die of this disease. (1) Randomized trials of mammography have demonstrated reductions in breast cancer mortality associated with screening from ages 50 to 74.(2–4) Trial results for women 40 to 49 years and 74 years and older were not conclusive and there were some problems in the design, conduct, and/or interpretation of the trials.(4,5) However, conducting additional trials to get more precise estimates of the mortality benefits from extending screening to women younger than 50 or older than 74 years or by testing different screening schedules is not feasible.

We developed models of breast cancer incidence and mortality in the US. These models are ideally suited for estimating the impact of screening under a variety of policies. (6;7) Modeling has the advantage of being able to hold selected conditions (e.g., screening intervals, test sensitivity) constant, facilitating comparison of strategies. Since all models make assumptions about unobservable events, use of several models provides a range of plausible effects and can illustrate the effects of differences in model assumptions.(7)

In this study, we use six established models to estimate the outcomes across 20 mammography screening strategies that vary by ages of initiation and cessation and screening interval among a cohort of US women. The results are intended to contribute to practice and guideline policy debates.

METHODS

The six models were developed independently within the National Cancer Institute's (NCI) Cancer Intervention and Surveillance Modeling Network ("CISNET")(7;8) and were exempt from IRB approval. The models have been described elsewhere. (7;9–15) Briefly, they share common features and inputs but differ in some ways (Appendix Table 1). Briefly, models from Erasmus MC (E), Georgetown (G), MD Anderson (M) and Wisconsin (W) include DCIS; models E and W specifically assume that some portion of DCIS are non-progressive and do not result in mortality; model W further assumes that some small invasive cancers are non-progressive. The Stanford (S) and Dana-Farber (D) models include only invasive cancer. Some groups model breast cancer in stages, but three (E, S and W) use tumor size and/or tumor growth. The models also differ in whether treatment affects the hazard of breast cancer death (G, S, D), results in a cure for some fraction (E, W) or both (M). Despite these differences, in previous collaborations all the models came to similar qualitative estimates of the relative contributions of screening and treatment to observed declines in breast cancer mortality. (7)

Model Overview

We use the six models to estimate the benefits, resource use (as measured by number of mammograms) and harms of 20 alternative screening strategies varying by starting and stopping age and interval (annual and biennial) (Table 1). The models begin with estimates

of what breast cancer incidence and mortality trends would have been in the absence of screening and treatment and then overlay screening use and improvements in survival associated with treatment. (7) We use a cohort of women born in 1960 beginning from age 25 and followed for their lifetimes. Generally breast cancer is depicted as having a preclinical screen detectable period (sojourn time) and a clinical detection point. Based on mammography sensitivity (or thresholds of detection), screening identifies disease in the preclinical screen-detection period and results in the identification of earlier stage/smaller cancers than might occur via clinical detection, resulting in breast cancer mortality reductions. Age, estrogen receptor (ER) status and tumor size/stage-specific treatment have independent impacts on mortality. Women can die of breast cancer or of other causes.

Model Data Parameters

All six modeling groups use a common set of age-specific parameters for breast cancer incidence, mammography test characteristics, treatment algorithms and treatment effects and non-breast cancer competing causes of death (Appendix Table 2). In addition to these common parameters, each model includes model-specific inputs (or intermediate outputs) to represent pre-clinical detectable times, lead time, dwell time within stages of disease and stage-distribution in unscreened vs. screened women based on their specific model structure. (7;9–15)

We use an age-period-cohort model to determine incidence for the cohort of women in the absence of screening. (16) This approach considers the impact of age, temporal trends in risk by cohort and time period in estimating what breast cancer incidence rates would have been without screening. Since we do not have data on future breast cancer incidence, we extrapolate forward assuming that future age-specific incidence rises as women age in the same manner as observed in 2000. To isolate the effect of technical effectiveness of screening and to assess the impact of screening on mortality holding treatment constant, models assume 100% compliance with screening and adherence to indicated treatment.

Three groups use the age-specific mammography sensitivity (and specificity) values observed in the Breast Cancer Surveillance Consortium (BCSC) program for detection of all breast cancers (invasive and in-situ combined). Separate values are used for initial and subsequent mammograms performed at either annual or biennial intervals. (17) Two of the models (D, G) use these data directly as an input parameter, (10;14) and one model (S) uses the data to calibrate the model. (13) The other three models (E, M, W) use the BCSC data as a guide and fit sensitivity estimates from this and other sources. (9;11;15)

All women who have estrogen receptor (ER) positive invasive tumors receive a hormonal (tamoxifen if <50 and anastrozole if >50+) and non-hormonal treatment with an anthracycline-based regimen. Women with ER negative invasive tumors receive non-hormonal therapy only. Women with DCIS who have ER positive tumors receive hormonal therapy only.(18) Treatment effectiveness is based on synthesis of recent clinical trials and modeled as a proportionate mortality risk reduction or the proportion cured. (19;20)

Benefits

We estimate the cumulative probability of dying of breast cancer from age 40 to death in the absence of screening. Screening benefit is then calculated as the percent breast cancer mortality reduction (vs. no screening). We also examine life years gained due to averted or delayed breast cancer death. Benefits are cumulated over the lifetime of the cohort to capture breast cancer mortality reductions (or life years gained) occurring years after the start of screening, after considering background non-breast cancer mortality.(21;22)

Harms

As measures of the burden that a regular screening program imposes on a population, we looked at three different potential screening "harms," including false positive mammograms, unnecessary biopsies and "over-diagnosis." We define the rate of false positive mammograms as the number of mammograms read as abnormal or needing further follow-up in women without cancer divided by the total number of all positive screening mammograms based on the specificity reported in the BCSC. (17) "Unnecessary" biopsies are defined post-hoc as the proportion of women with false positive screens who receive a biopsy.(23) "Over-diagnosis" is defined as the proportion of cases in each strategy that would not have clinically surfaced in a woman's life time (due to lack of progressive potential or competing morality) among all cases arising from age 40 onward.

Analysis

Model results for the 20 strategies are compared to select the most "efficient" approach among competing strategies. In decision analysis, a new intervention is considered more "efficient" than a comparison intervention if it results in gains in health outcomes, such has life years or deaths averted, while consuming fewer resources (or costs) than the comparator. If the new intervention results in worse outcomes and requires a greater investment, it is "inefficient", and would not be considered for further use. In economic analysis, inefficient strategies are said to be "dominated" when this occurs. To rank the screening strategies we first look at the results of each model independently. For a particular model, a strategy that requires more mammograms (our measure of resource use) but has lower relative percent mortality reduction (or life years gained) is considered "inefficient" or "dominated" by other strategies. To evaluate strategies given results from all six models together, we classify strategies in the following way: If a strategy is dominated in all or 5 of 6 of the models, it is considered dominated overall. If a screening strategy is not dominated in any of the models, it is classified as efficient. For strategies with mixed results across the models, we classify the strategy as borderline.

After eliminating all dominated strategies, the remaining strategies are represented as points on a graph that plots the average number of mammograms versus the percent mortality decline (or years of life gained) for each individual model. The efficiency frontier for each graph is obtained by identifying the sequence of points that represent the largest incremental gain in percent mortality reduction (or gain in life years) per additional screening mammogram. Screening strategies that fall on this frontier are the most efficient (i.e., no alternative exists that provides more benefit for fewer mammograms performed).

Sensitivity Analysis

We conduct a sensitivity analysis to see if our conclusions about the ranking of strategies changes when we vary input parameters. First, we investigate the impact of assuming that mammography sensitivity for a given age, screening round and screening interval is 10 percentage points lower than observed. Second, we examine whether ranking of strategies varies if treatment includes newer hormonal and non-hormonal adjuvant regimens (e.g., taxanes, etc). Third, since adjuvant therapy is unlikely to reach 100% of women as modeled in our base analysis, we re-assess the ranking of strategies if we assume that actual observed current treatment patterns apply to the cohort. (24)

Model Validation and Uncertainty

Each of the models has a different structure, assumptions, and some varying input parameters, so there is no single method that can be used to validate results against some external gold standard. For instance, since some models used results from screening trials

(or SEER data) for calibration or as input parameters, we can not use comparisons of projected mortality reductions to trial results to validate all of the models. In addition, we can not directly compare the results of this analysis, which uses 100% actual screening for all women at specified intervals to screening trial results where invitation to screening and participation was variable. In our prior work, results of each model accurately projected independently estimated trends in the absence of intervention and closely approximated modern stage distributions and observed mortality trends. (7;9–11;13–15) Overall, using six models to project a range of plausible screening outcomes provides implicit cross-validation, with the range of results from the models as a measure of uncertainty.

Role of the Funding Agencies

This work was done under contracts from the Agency for Healthcare Research and Quality (AHRQ) and NCI and grants from the NCI. The NCI provided some data and technical assistance and AHRQ reviewed the manuscript. Model results are the sole responsibility of the investigators.

RESULTS

In the absence of screening, the models predict a cumulative probability of developing breast cancer over a woman's lifetime starting at age 40 ranging from 12% to 15%. Without screening the median probability of dying of breast cancer after age 40 is 3.0% across the six models. Thus, if a particular screening strategy leads to a 10% breast cancer mortality reduction, then the probability of breast cancer death would be reduced from 3.0% to 2.7%, or 3 deaths averted per 1000 women screened.

Benefits

The six models produce consistent results on the ranking of the strategies (Table 2). There are 8 approaches that are "efficient" in all models (i.e., not dominated since they provide additional mortality reductions for added use of mammograms); 7 of these 8 have a biennial interval and all but 2 start at age 50. Figure 1 depicts these results graphically and again we see that the overwhelming majority of strategies on the efficiency frontier have a biennial interval. Screening every other year from ages 50 to 69 is an efficient strategy for reducing breast cancer mortality in all models. In all of the models, biennial screening starting at age 50 and continuing up through ages 74, 79 or 84 years are of fairly similar efficiency.

Examining benefits in terms of life years gained (Appendix Table 1), 6 of the 8 consistently non-dominated strategies have a biennial interval. In contrast to results for mortality reduction, half of the non-dominated strategies include screening initiation at age 40. Annual screening strategies that include screening until age 79 or 84 are on the efficiency frontier (Appendix Figure 2), but are less resource- efficient than biennial approaches for increasing life years gained.

As another way to examine the effect of screening interval, we calculated the proportion of the annual benefit (in terms of mortality reduction) that could be achieved by screening biennially for each screening strategy and model (Table 3). Screening biennially maintains an average of 81% (range across strategies and models 67–99%) of the benefits achieved by annual screening.

We also examined the incremental benefits gained by extending screening from ages 50 to 69 to either earlier or later ages of initiation and cessation (Tables 4a and b). Continuing screening to age 79 (vs. 69) results in a median increase in percent mortality reduction of 8% (range 7–11%) and 7% (6–10%) under annual and biennial intervals, respectively. If screening begins at age 40 (vs. 50) and continues to 69, all models project additional albeit

small reductions in breast cancer mortality (median 3% mortality reduction with either annual or biennial intervals) (Table 4a). This translates into a median of one additional breast cancer death averted (range 1 to 2) per 1000 women under a strategy of annual screening from age 40 to 69 (vs. 50 to 69). Thus, there are greater mortality reductions that could be achieved by having an older age of cessation than by initiating screening at an earlier age.

However, when life years gained is the outcome measure 3 of the models conclude that there are greater benefits from extending screening to the younger rather than the older age group (Table 4b). For instance, starting annual screening at age 40 (vs. 50) and continuing annually to 69 yields a median of 33 (range 11–58) additional life years per 1000 women screened while extending annual screening to age 79 (vs. 69) yields only a median of 24 (range 18–38) added life years per 1000 women screened.

Harms

The models all project similar rates of false positive mammograms over the lifetime of screened women across the screening strategies (Tables 5a and b). There are more false positive screens in strategies that included screening from ages 40 to 49 than those that initiated screening at age 50 or later and more with annual strategies than biennial strategies. For instance, annual screening from ages 40 to 69 yields 2250 false positive tests for every 1000 women screened over this time period, almost double the number of false positives when screening biennially in this age group. The proportion of biopsies that occur as a result of these false positive screens that are retrospectively deemed unnecessary (i.e., the woman did not have cancer) is about 7%, so that substantially more women will undergo needless biopsies under annual than biennial schedules.

Five of the six models estimated rates of over-diagnosis and all of these show an increase in the risk of over-diagnosis as age increases (data not shown). Although the increase with age occurs over the entire age range considered in the different screening strategies, the rate of increase accelerates in the older age groups, largely due to increasing rates of competing causes of mortality in older age groups. Rates of over-diagnosis were higher for DCIS than for invasive disease, proportionately affecting younger women more since the percent of cases that are diagnosed as DCIS is larger at younger ages. Overall, however, initiating screening at age 40 (vs. 50) had a smaller impact on over-diagnosis than extending screening beyond age 69. Biennial strategies decrease the rate of over-diagnosis, but by a factor much less than one half. The absolute estimate of over-diagnosis varied between models depending on whether or not DCIS was included and the assumptions related to progression of DCIS and invasive disease, reflecting the uncertainly in the current knowledge base.

Sensitivity Analysis

The overall conclusions were robust across the six models under different assumptions about mammography sensitivity, treatment patterns and treatment effectiveness (not shown).

DISCUSSION

This study uses six established models that employ common inputs but differing approaches and assumptions to extend previous randomized mammography screening trials results to the US population and to age groups where trial results are less conclusive. All six modeling groups concluded that the most efficient screening strategies are those that include a biennial screening interval. Conclusions about the optimal starting ages for screening depend more on the metric chosen for evaluating outcomes. If the goal of a national screening program is to reduce mortality in the most efficient manner, then programs that screen biennially from

ages 50 to either age 69, 74 or 79 are among the most efficient on the basis of the ratio of benefits to the number of screening examinations. If the goal of a screening program is to efficiently maximize the number of life years gained, then the preferred strategy would be to screen biennially starting at age 40. Decisions about the best starting and stopping ages also depend on tolerance for false positive screens and rates of over-diagnosis.

The conclusion of this modeling analysis that biennial intervals are more efficient and provide a better balance of benefits and harms than annual intervals is contrary to some US current practices. (25–27) However, our result that biennial screening is more efficient than annual screening is consistent with prior modeling research (28–32) and screening trials, most of which used 2 year intervals. (2–5) The model results are also congruent with reports showing similar intermediate cancer outcomes (e.g., stage distribution) between programs using annual and biennial screening, especially among women 50 and older. (33–37) In addition, we demonstrated substantial increases in the numbers of false positive screens and unnecessary biopsies associated with annual intervals and these harms are reduced by almost 50% with biennial intervals. Our results are also consistent with what we know about disease biology. In the majority of women most tumors are slow growing and this proportion increases with age, (38) so that there is little loss in survival benefit for screening every year vs. every other year. Conversely, for women with aggressive, faster growing tumors even annual screening is not likely to confer a survival advantage. Guidelines in other countries include biennial screening. (4) However, whether it will be practical or acceptable to change the existing US practice of annual screening can not be addressed by our models.

In all the models, some breast cancer mortality reduction, albeit small, was seen with strategies that started screening at age 40 vs. age 50. By being able to model millions of observations, models are well suited to detecting small differences in a group over time that might not be seen in even the largest clinical trial with a 10–15 year follow-up period. (4;39–42) If program benefits are to be measured in life years, the metric most commonly used in cost-effectiveness analysis, then our results suggest that initiating screening at age 40 saves more life years than extending screening past age 69 (albeit at the cost of an increase in the number of false positive mammograms).

Historically, breast cancer screening recommendations have suggested an upper age limit for screening cessation based on considerations of decreasing program efficiency due to competing mortality. (26;43) Our result that screening strategies that include an upper age limit beyond age 69 remain on the efficiency frontier (albeit with low incremental gains over strategies that stop screening at earlier ages and with greater harms) is consistent with previously reported results of screening benefit from observational and modeled data. (31;32;44–47) However, the observational data reports may have been confounded by inability to capture lead time and length biases.(48–50) Any benefits of screening older women must be balanced against possible harms. For instance, the probability of over-diagnosis increases with age and increases more dramatically for the oldest ages. Model estimates for the oldest ages also have more uncertainty compared to estimates for age groups 50 to 74 due to the paucity of primary data on breast cancer natural history and the absence of screening trial data after age 74. With the demographic pressure of an aging society, more research will be needed to fully understand the natural history of disease and the balance of risks and benefits of treatment in the older age groups.(38;50)

The results of this modeling analysis also highlight the need for better primary data on the natural history of DCIS and small invasive cancers to draw reliable conclusions on the absolute magnitude of over diagnosis associated with different screening schedules. (37;51) Clinical investigation, (52) follow-up in screening trials,(53)epidemiologic trends in

incidence (54) and prior modeling efforts (9;55) all indicated that some fraction of DCIS cases will not progress, (56;57) but that fraction is not known.

The collaboration of six groups with differing modeling philosophies and approaches to estimate the same endpoints using a common set of data provides an excellent opportunity to cross-replicate data generated from modeling, represent uncertainty related to modeling assumptions and structure and give insight into which results are consistent across modeling approaches and which are dependent on model assumptions. The resulting conclusions about the ranking of screening strategies were very robust and should provide greater credibility than inferences based on one model alone.

Despite our consistent results, there are several caveats that should be considered in evaluating our results.(58) First, our models provide estimates of the average benefits and harms expected across a cohort of women and do not reflect personal data for individual women. Also, while our models project mortality reductions similar to those observed in clinical trials, the range of results includes higher mortality reductions than achieved in the trials since we model lifetime screening and assume compliance with all screening and treatment. The trials followed women for limited numbers of years and have some non-adherence. The models also do not capture differences in outcomes among certain risk sub-groups, such as women with BRCA 1 or 2 genetic susceptibility mutations, those who are healthier or sicker than average or African-American women who appear to have more disease at younger ages than Whites. (59) Second, the outcomes considered do not capture morbidity associated with surgery for screen detected disease (60) or decrements in quality of life associated with false positive screening or living with earlier knowledge of a cancer diagnosis or over-diagnosis.(61)

Third, in estimating lifetime results we projected breast cancer trends from background incidence rates of a 1960 birth cohort extrapolated forward in time. However, future background incidence (and mortality) may change as the result of several different forces, such as changes in patterns of reproduction, lower use of hormone replacement therapy after the year 2002 or prescription of tamoxifen or other agents for primary disease prevention, increasing rates of obesity and further advances in treatment (e.g., trastuzumab) (62) While most models portray known differences in biology by age (e.g., distribution of ER positive tumors, sensitivity of screening and length of the pre-clinical sojourn times), some aspects of the natural history of disease are not known and/or can not be fully captured.

We assumed 100% compliance with screening and treatment to evaluate program efficacy. Benefits will always fall short of the projected results since adherence is not perfect. If actual compliance varies systematically by age or other factors, it is possible that the ranking of strategies could change. In addition, we did not consider “mixed” strategies (e.g., screening annually from ages 40 to 49 and then biennially from age 50 to 79) as was done in some trials (5) and other analyses (36; 63) for several reasons. First, we found that the benefits of screening from 40 to 49 were small. Benefits in this age group were also associated with harms in terms of false positive tests and unnecessary biopsies. Thus, while strategies that include annual screening from 40–49 might be efficient, this would be largely driven by the more favorable balance of benefits and harms after age 50. Finally, we judged that mixed strategies are very difficult to communicate to consumers and implement in public health practice. Finally, we did not discount benefits or include costs in our analysis although the average number of mammograms per woman strategy (and inclusion of false positive screens) provides some proxy of resource consumption. Even with these acknowledged limits the models demonstrate meaningful, qualitatively similar outcomes despite variations in structure and assumptions.

Overall, the evaluation of screening strategies by the six models suggests that optimal program design would be based on biennial intervals. Choices about optimal ages of initiation and cessation will ultimately depend on program goals, resources, weight attached to the presence of trial data, the balance of harms and benefits and considerations of efficiency and equity.

Acknowledgments

Supported by National Cancer Institute Cooperative Agreements #2U01CA088270 (Lee), 2U01CA088283 (Mandelblatt, DeKoning, Berry), 2U01CA088248 (Plevritis), and F32 CA125984 (Stout). Portions of this work were performed under contract #HHSN261200800769P (Mandelblatt).

We thank the Breast Cancer Surveillance Consortium (BCSC) investigators, participating mammography facilities, and radiologists for the data they provided to the BCSC that were used to inform some of our model data input parameters. Breast Cancer Surveillance Consortium data collection was supported by NCI-funded Breast Cancer Surveillance Consortium co-operative agreements (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040) and several state public health departments and cancer registries throughout the U.S. For a full description of the cancer registry support, please see: http://breastscreening.cancer.gov/work/acknowledgement.html. A list of the BCSC investigators and procedures for requesting BCSC data for research purposes is provided at: http://breastscreening.cancer.gov/.

To Mary Barton, MD, MPP and William Lawrence, MD, MSc from the Agency for Healthcare Research and Quality, members of the US Preventive Services Task Force, the Oregon Evidence Based Practice Center, Ann Zauber, PhD and Karla Kerlikowske, MD for helpful comments and review of earlier versions of this paper.

To Cornerstone Systems Northwest, Inc. (Lauren Clarke) for data management and web site support (NCI contract HHSN261200800002C).

To Jackie Ford and Aimee Near for manuscript preparation.

References

1. Jemal A, Siegel R, Ward Em Hao Y, Xu J, thus MJ. Cancer Statistics 2009. American Cancer Society CA Cancer. J Clin. 2009; 59:225–249.
2. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet. 2002; 359:909–19. [PubMed: 11918907]
3. Tabar L, Vitak B, Chen H, Duffy S, Yen MF, Chiang CG, Krusemo UB, Tot T, Smith RA. The Swedish Two-County trial twenty years later: Updated mortality results and new insights from long-term follow-up. Radiologic Clinics of North America. 2000; 38 (4):625–51. [PubMed: 10943268]
4. International Agency for Research on Cancer. Breast Cancer Screening. In: Vainio, H.; Bianchini, F., editors. IARC Handbook on Cancer Prevention, Report No 7. Lyon: IARC; 2002.
5. 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; 368:2053–60. [PubMed: 17161727]
6. Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR. Assessing the effectiveness of health interventions for cost-effectiveness analysis. Panel on Cost-Effectiveness in Health and Medicine. J Gen Intern Med. 1997; 12:551–58. [PubMed: 9294789]
7. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005; 353:1784–92. [PubMed: 16251534]
8. CISNET. [Accessed September 15. 2008.] cisnet.cancer.gov/profiler/
9. 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. 2006:37–47. [PubMed: 17032893]
10. 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. 2006:47–55. [PubMed: 17032894]

11. 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. 2006:30–36. [PubMed: 17032892]
12. 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. 2006:96–105. [PubMed: 17032899]
13. 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. 2006:86–95. [PubMed: 17032898]
14. Lee S, Zelen M. A stochastic model for predicting the mortality of breast cancer. J Natl Cancer Inst Monogr. 2006:79–86. [PubMed: 17032897]
15. 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. 2006:56–65. [PubMed: 17032895]
16. Holford TR, Cronin KA, Mariotto AB, Feuer EJ. Changing patterns in breast cancer incidence trends. J Natl Cancer Inst Monogr. 2006:19–25. [PubMed: 17032890]
17. Breast Cancer Surveillance Consortium. [accessed April 11. 2008.] Performance Measures for 3,603,832 Screening Mammography Examinations from 1996 to 2006 by Age & Time (Months) Since Previous Mammography. http://breastscreening.cancer.gov/data/performance/perf_age_time.html
18. National Comprehensive Cancer Network. NCCN Clinical Practice guidelines in oncology. 2008; 2 http://www.nccn.org/professionals/physician_gls/f_guidelines.asp.
19. Clarke M, Coates AS, Darby SC, Davies C, Gelber RD, Godwin J, et al. Adjuvant chemotherapy in oestrogen-receptor-poor breast cancer: patient-level meta-analysis of randomised trials. Lancet. 2008; 371:29–40. [PubMed: 18177773]
20. Early Breast Cancer Trialists' Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005; 365:1687–717. [PubMed: 15894097]
21. Rosenberg MA. Competing risks to breast cancer mortality. J Natl Cancer Inst Monogr. 2006:15–19. [PubMed: 17032889]
22. 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. 2006:112–21. [PubMed: 17032901]
23. Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM, et al. Performance benchmarks for screening mammography. Radiology. 2006; 241:55–66. [PubMed: 16990671]
24. 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. 2006:7–15. [PubMed: 17032888]
25. Smith RA, Saslow D, Sawyer KA, Burke W, Costanza ME, Evans WP III, et al. American Cancer Society guidelines for breast cancer screening: update 2003. CA Cancer J Clin. 2003; 53:141–69. [PubMed: 12809408]
26. National Cancer Institute. [accessed September 15. 2008.] http://www.cancer.gov/newscenter/mammstatement31jan02
27. [accessed September 15. 2008.] Medicare. http://www.medicare.gov/Health/Mammography.asp
28. Salzmann P, Kerlikowske K, Phillips K. Cost-effectiveness of extending screening mammography guidelines to include women 40 to 49 years of age. Ann Intern Med. 1997; 127:955–65. [PubMed: 9412300]
29. 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; 98:774–82. [PubMed: 16757702]
30. Lee S, Huang H, Zelen M. Early detection of disease and scheduling of screening examinations. Stat Methods Med Res. 2004; 13:443–56. [PubMed: 15587433]

31. Mandelblatt JS, Schechter CB, Yabroff KR, Lawrence W, Dignam J, Extermann M, et al. Toward optimal screening strategies for older women. Costs, benefits, and harms of breast cancer screening by age, biology, and health status. J Gen Intern Med. 2005; 20:487–96. [PubMed: 15987322]
32. Kerlikowske K, Salzmann P, Phillips KA, Cauley JA, Cummings SR. Continuing screening mammography in women aged 70 to 79 years: impact on life expectancy and cost-effectiveness. JAMA. 1999; 282:2156–63. [PubMed: 10591338]
33. Hofvind S, Vacek PM, Skelly J, Weaver DL, Geller BM. Comparing screening mammography for early breast cancer detection in Vermont and Norway. J Natl Cancer Inst. 2008; 100:1082–91. [PubMed: 18664650]
34. Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, et al. Comparison of screening mammography in the United States and the United Kingdom. JAMA. 2003; 290:2129–37. [PubMed: 14570948]
35. Smith-Bindman R, Ballard-Barbash R, Miglioretti DL, Patnick J, Kerlikowske K. Comparing the performance of mammography screening in the USA and the UK. J Med Screen. 2005; 12:50–54. [PubMed: 15814020]
36. White E, Miglioretti DL, Yankaskas BC, Geller BM, Rosenberg RD, Kerlikowske K, et al. Biennial versus annual mammography and the risk of late-stage breast cancer. J Natl Cancer Inst. 2004; 96:1832–39. [PubMed: 15601639]
37. Wai ES, D'yachkova Y, Olivotto IA, Tyldesley S, Phillips N, Warren LJ, et al. Comparison of 1- and 2-year screening intervals for women undergoing screening mammography. Br J Cancer. 2005; 92:961–66. [PubMed: 15714210]
38. Fracheboud J, Groenewoud JH, Boer R, Draisma G, de Bruijn AE, Verbeek AL, de Koning HJ. Seventy-five years is an appropriate upper age limit for population-based mammography screening. Int J Cancer. 2006 Apr 15; 118(8):2020–5. [PubMed: 16287064]
39. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. Ann Intern Med. 2002; 137:305–12. [PubMed: 12204013]
40. Elmore JG, Armstrong K, Lehman CD, Fletcher SW. Screening for breast cancer. JAMA. 2005; 293:1245–56. [PubMed: 15755947]
41. Elmore JG, Reisch LM, Barton MB, Barlow WE, Rolnick S, Harris EL, et al. Efficacy of breast cancer screening in the community according to risk level. J Natl Cancer Inst. 2005; 97:1035–43. [PubMed: 16030301]
42. Norman SA, Russell LA, Weber AL, Coates RJ, Zhou L, Bernstein L, et al. Protection of mammography screening against death from breast cancer in women aged 40–64 years. Cancer Causes Control. 2007; 18:909–18. [PubMed: 17665313]
43. U.S.Preventive Services Task Force. Screening for Breast Cancer. USPSTF; 2002. http://www.ahrq.gov/clinic/uspstf/uspsbrca.htm
44. McCarthy EP, Burns RB, Freund KM, Ash AS, Shwartz M, Marwill SL, et al. Mammography use, breast cancer stage at diagnosis, and survival among older women. J Am Geriatr Soc. 2000; 48:1226–33. [PubMed: 11037009]
45. Lash TL, Fox MP, Buist DS, Wei F, Field TS, Frost FJ, et al. Mammography surveillance and mortality in older breast cancer survivors. J Clin Oncol. 2007; 25:3001–6. [PubMed: 17548838]
46. Badgwell BD, Giordano SH, Duan ZZ, Fang S, Bedrosian I, Kuerer HM, et al. Mammography before diagnosis among women age 80 years and older with breast cancer. J Clin Oncol. 2008; 26:2482–88. [PubMed: 18427152]
47. Boer R, de Koning HJ, van Oortmarssen GJ, van der Maas PJ. In search of the best upper age limit for breast cancer screening. Eur J Cancer. 1995; 31A:2040–2043. [PubMed: 8562162]
48. Berry D, Baines C, Baum M, Dickersin K, Fletcher S, et al. Flawed inferences about screenng mammography's benefit based on observational data. J Clin Oncol. 2008 in press.
49. Schonberg MA, McCarthy EP. Mammography screening among women aged 80 and older: consider the risks. J Clin Oncol. 2008 in press.

50. Mandelblatt J, Silliman R. Hanging in the balance: making decisions about the benefits and harms of breast cancer screening among the oldest old without a safety net of scientific evidence. J Clin Oncol. 2008; 26 n press.
51. Bryan BB, Schnitt SJ, Collins LC. Ductal carcinoma in situ with basal-like phenotype: a possible precursor to invasive basal-like breast cancer. Mod Pathol. 2006; 19:617–21. [PubMed: 16528377]
52. Kerlikowske K, Molinaro A, Cha I, Ljung BM, Ernster VL, Stewart K, et al. Characteristics associated with recurrence among women with ductal carcinoma in situ treated by lumpectomy. J Natl Cancer Inst. 2003; 95:1692–702. [PubMed: 14625260]
53. Moss S. Overdiagnosis and overtreatment of breast cancer: overdiagnosis in randomised controlled trials of breast cancer screening. Breast Cancer Res. 2005; 7:230–234. [PubMed: 16168145]
54. Feuer EJ, Etzioni R, Cronin KA, Mariotto A. The use of modeling to understand the impact of screening on U.S. mortality: examples from mammography and PSA testing. Stat Methods Med Res. 2004; 13:421–42. [PubMed: 15587432]
55. de Koning HJ, Draisma G, Fracheboud J, de Bruijn A. Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data. Breast Cancer Res. 2006; 8:202. [PubMed: 16524452]
56. Burstein HJ, Polyak K, Wong JS, Lester SC, Kaelin CM. Ductal carcinoma in situ of the breast. N Engl J Med. 2004; 350:1430–1441. [PubMed: 15070793]
57. Jones JL. Overdiagnosis and overtreatment of breast cancer: progression of ductal carcinoma in situ: the pathological perspective. Breast Cancer Res. 2006; 8:204. [PubMed: 16677423]
58. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value Health. 2003; 6:9–17. [PubMed: 12535234]
59. Mandelblatt, JS.; Liang, W.; Sheppard, VB.; Wang, J.; Isaacs, C. Breast Cancer in Minority Women. 4. Harris, J.; Lippman, M., editors. Philadelphia: Lippincott-Raven; 2008. in press
60. El Tamer MB, Ward BM, Schifftner T, Neumayer L, Khuri S, Henderson W. Morbidity and mortality following breast cancer surgery in women: national benchmarks for standards of care. Ann Surg. 2007; 245:665–71. [PubMed: 17457156]
61. Bonomi AE, Boudreau DM, Fishman PA, Ludman E, Mohelnitzky A, Cannon EA, et al. Quality of life valuations of mammography screening. Qual Life Res. 2008; 17:801–14. [PubMed: 18491217]
62. Ravdin PM, Cronin KA, Howlader N, et al. The decrease in breast-cancer incidence in 2003 in the United States. N Engl J Med. 2007; 356:1670–1674. [PubMed: 17442911]
63. Buist DS, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40 to 49 years. JNCI. 2004; 96(19):1432–40. [PubMed: 15467032]

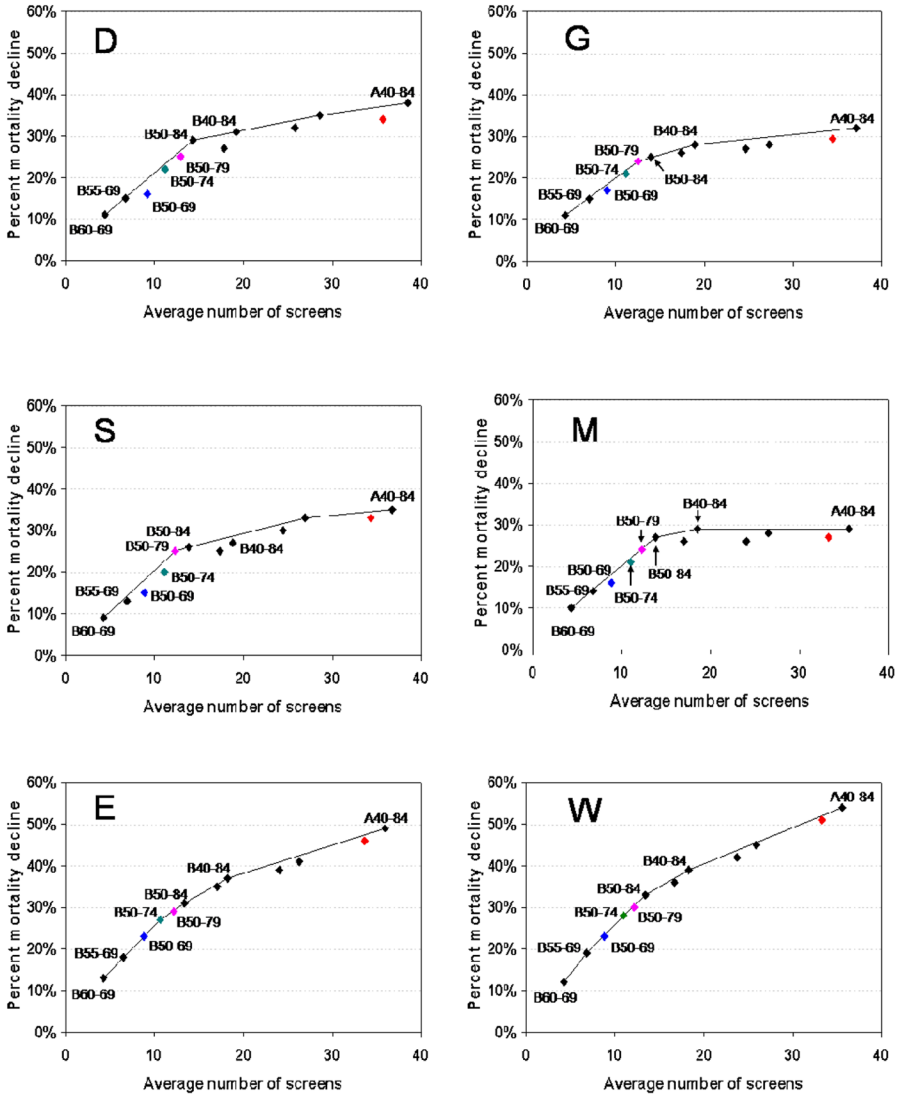
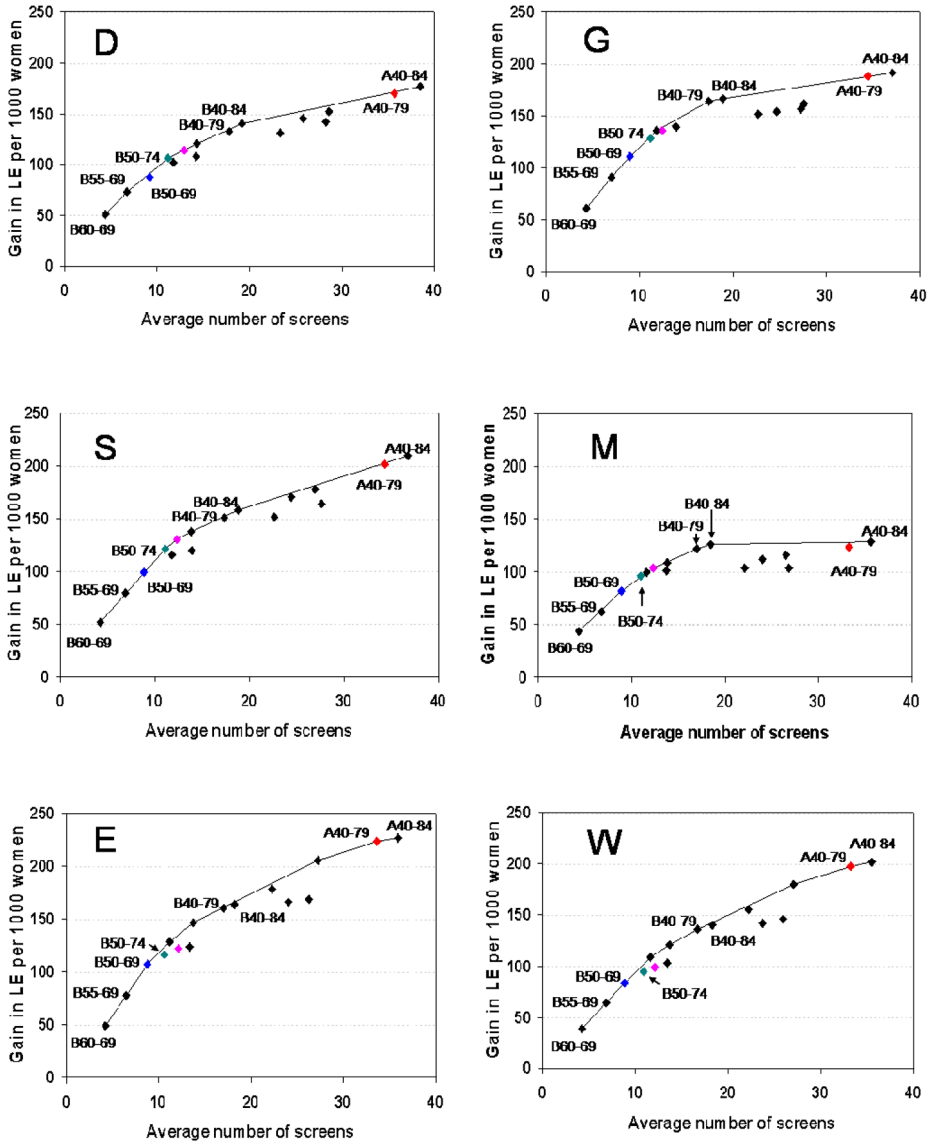


Figure 1. Percent Breast Cancer Mortality Reduction vs. Number of Mammography Screens per Woman by Model and Screening Strategy

The panels in this figure show an efficiency frontier graph for each model. The graph plots the average number of mammograms performed per women against the percent mortality reduction for each screening strategy (vs. no screening). We plot efficient strategies (i.e., those where increases in use of mammography resources result in greater mortality reduction than the next least intensive strategy) in all six models. We also plot “borderline” strategies (approaches that are efficient in some models but not in others). The line between strategies that is drawn represents the “efficiency frontier”. Strategies on this line would be considered efficient in that they achieve the greatest gain per use of mammography resources compared to 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 prior strategies and could indicate a point at which additional investment (use of screening) might be considered as having a low return (benefit).

To highlight efficient strategies that decision makers might want to consider, we have color coded the strategies that might be considered most efficient overall across the models. We also highlight one common current approach (annual screening 40–79), although it is below the efficiency frontier in most models. Blue represents biennial screening from age 50–69
Green represents biennial screening from age 50–74
Pink represents biennial screening from age 50–79
Red represents annual screening from age 40–79
Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G (Georgetown U.), M (M.D. Anderson Cancer Center), S (Stanford U.), W (U. of Wisconsin/Harvard)



Appendix Figure 1. Life Years Gained vs. Number of Mammography Screens per Woman by Model and Screening Strategy

The panels in this figure show an efficiency frontier graph for each model. The graph plots the average number of mammograms performed per women against the percent mortality reduction for each screening strategy (vs. no screening). We plot efficient strategies (i.e., those where increases in use of mammography resources result in greater mortality reduction than the next least intensive strategy) in all six models. We also plot “borderline” strategies (approaches that are efficient in some models but not in others). The line between strategies that is drawn represents the “efficiency frontier”. Strategies on this line would be considered efficient in that they achieve the greatest gain per use of mammography resources compared to 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 prior strategies and could indicate a point at which additional investment (use of screening) might be considered as having a low return (benefit).

To highlight efficient strategies that decision makers might want to consider, we have color coded the strategies that might be considered most efficient overall across the models. We also highlight one common current approach (annual screening 40–79), although it is below the efficiency frontier in most models.
Blue represents biennial screening from age 50–69
Green represents biennial screening from age 50–74
Pink represents biennial screening from age 50 to 79
Red represents annual screening from age 40 to 79
Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G (Georgetown U.), M (M.D. Anderson Cancer Center), S (Stanford U.), W (U. of Wisconsin/Harvard)

Table 1

Breast Cancer Screening Strategies

- No screening
- Screen from ages 40–69
- Screen from ages 40–79
- Screen from ages 40–84
- Screen from ages 45–69
- Screen from ages 50–69
- Screen from ages 50–74
- Screen from ages 50–79
- Screen from ages 50–84
- Screen from ages 55–69
- Screen from ages 60–69

Each strategy was evaluated using an annual or biennial schedule for a total of 20 strategies; we include no screening for comparisons.

Table 2

The Average Number of Screening Exams per 1000 Women and the Percent Breast Cancer Mortality Decline among Screened Women for each Model by Screening Strategy

Screening Strategies	Average # of screens per 1000women[1]	Percent Breast Cancer Mortality Reduction (vs. no screening) Models[2]					
		D	E	G	M	S	W
Efficient Strategies (not dominated in 6 of 6 models)							
B 60–69	4263	11%	13%	11%	10%	9%	12%
B 55–69	6890	15%	18%	15%	14%	13%	19%
B 50–69	8947	16%	23%	17%	16%	15%	23%
B 50–74	11066	22%	27%	21%	21%	20%	28%
B 50–79	12366	25%	29%	24%	24%	25%	30%
B 50–84	13837	29%	31%	25%	27%	26%	33%
B 40–84	18708	31%	37%	28%	29%	27%	39%
A 40–84	36550	38%	49%	32%	29% **	35%	54%
Borderline Strategies (dominated in 2–3 of 6 models)							
B 40–79	17241	27%[3]	35%	26%	26%	25%	36%
A 50–79	24419	32%	39%	27%	26%	30%	42%
A 50–84	26905	35%	41%	28%	28%	33%	45%
A 40–79	34078	34%	46%	30%	27%	33%	51%
Inefficient/Dominated Strategies (dominated in all 6 models)							
A 60–69	8438	14%	18%	13%	12%	12%	17%
B 45–69	11694	18%	26%	20%	19%	17%	27%
A 55–69	13009	18%	25%	17%	15%	16%	26%
B 40–69	13831	18%	28%	20%	19%	16%	29%
A 50–69	17733	21%	31%	20%	18%	20%	33%
A 50–74	21330	27%	35%	24%	22%	26%	38%
A 45–69	22546	23%	35%	22%	20%	22%	39%
A 40–69	27428	24%	39%	23%	20%	22%	43%

A=Annual

B=Biennial

[1]Average number of mammograms across models. Not all possible mammograms in the age interval are obtained in strategies that continue to the oldest age groups since many women die as the result of other causes before screening would occur.

[2]Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G (Georgetown U.), M (M.D. Anderson Cancer Center), S (Stanford U.), W (University of Wisconsin/Harvard)

[3]Shaded areas in the table show strategies that are dominated (“inefficient”) within a specific model; a strategy is classified as dominated if there is another strategy (from either the Efficient, Borderline or Inefficient/Dominated categories) that results in an equal or higher percent mortality decline with fewer average screening exams.

** Due to rounding, this strategy appears to be dominated, but the actual result is 29.4%

Table 3

Percent Breast Cancer Mortality Reduction Maintained when Moving from an Annual Screening Interval to a Biennial Interval by Screening Strategy and Model

Model	Percent Mortality Reduction Maintained by Screening Strategy									
	50–69	**40–69**	**45–69**	**40–79**	**40–84**	**55–69**	**60–69**	**50–74**	**50–79**	**50–84**
D	76%	75%	78%	79%	82%	83%	79%	81%	78%	83%
E	75	73	74	75	75	75	73	76	75	76
G	85	86	91	87	88	91	86	89	88	89
M	90	96	97	97	99	92	84	95	93	95
S	74	73	78	76	77	80	74	79	85	79
W	68	67	70	70	71	71	70	72	70	73

Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G (Georgetown U.), M (M.D. Anderson Cancer Center), S (Stanford U.), W (U. of Wisconsin/Harvard)

Differences in the range of results reflect differences in modeling approaches. For example, the benefit of screening in model M is modeled through stage shift, as with most other models, but also includes a “beyond stage shift” factor based on a cure fraction for small tumors. However, since many of these “cures” occur among women with invasive cancers that are not lethal, finding such cancers a year earlier confers very little mortality advantage to annual (vs. biennial) screening.

Table 4a

Incremental Change in Percent Breast Cancer Mortality Reduction by Age of Screening Initiation and Cessation

	Start Age 40 (vs. 50)*				Stop Age 79 (vs. 69)**			
	Difference in % breast cancer mortality reduction		Number of breast cancer deaths averted/1000 women		Difference in % breast cancer mortality reduction		Number of breast cancer deaths averted/1000 women	
Model	**Annual**	**Biennial**	**Annual**	**Biennial**	**Annual**	**Biennial**	**Annual**	**Biennial**
D	3%	2%	1	1	11%	9%	3	3
E	8%	5%	2	1	8%	6%	2	2
G	3%	3%	1	1	7%	7%	2	2
M	2%	3%	1	1	7%	7%	2	2
S	2%	1%	1	1	10%	10%	4	4
W	10%	6%	2	1	8%	6%	2	1
Median across models	3%	3%	1	1	8%	7%	2	2

*Incremental difference between screening from 40–69 vs. 50–69

**Incremental difference between screening from 50–79 vs. 50–69

Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G (Georgetown U.), M (M.D. Anderson Cancer Center), S (Stanford U.), W (U. of Wisconsin/Harvard)

Table 4b

Incremental Change in Life Years Gained per 1000 Women by Age of Screening Initiation and Cessation

	Start Age 40 (vs. 50)*		Stop Age 79 (vs. 69)**	
	Difference in Life Years Gained/1000 women		Difference in Life Years Gained/1000 women	
Model	**Annual**	**Biennial**	**Annual**	**Biennial**
D	25	20	28	26
E	58	40	18	15
G	34	29	27	25
M	11	18	21	21
S	32	21	38	31
W	57	37	19	15
Median across models	33	25	24	23.5

*Incremental difference between screening from 40–69 vs. 50–69

**Incremental difference between screening from 50–79 vs. 50–69

Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G (Georgetown U.), M (M.D. Anderson Cancer Center), S (Stanford U.), W (U. of Wisconsin/Harvard)

Table 5a

Summary of Benefits and Harms-Comparison of Different Starting Ages Using Exemplar Model [1]

		Potential Benefits (vs. no screening)			Potential Harms[2]	
Strategy	**Average Screens/1000**	**% Mortality Reduction**	**Cancer Deaths Averted/1000**	**Life Years Gained per 1000**	**# False positives/1000**	**# of unnecessary biopsies/1000**
Biennial						
B 40–69	13865	16%[3]	6.1	120	1250	88
B 45–69	11771	17%	6.2	116	1050	74
B 50–69	8944	15%	5.4	99	780	55
B 55–69	6941	13%	4.9	80	590	41
B 60–69	4246	9%	3.4	52	340	24
Annual						
A 40–69	27583	22%	8.3	164	2250	158
A 45–69	22623	22%	8.0	152	1800	126
A 50–69	17759	20%	7.3	132	1350	95
A 55–69	13003	16%	6.1	102	950	67
A 60–69	8406	12%	4.6	69	600	42

A= Annual

B= Biennial

[1] Results are from Model S. Model S was chosen as an exemplar model to summarize the balance of benefits and harms associated with screening 1000 women under a particular screening strategy.

[2] Over-diagnosis is another significant harm associated with screening. However, given the uncertainty in the knowledge base about DCIS and small invasive tumors, we felt that the absolute estimates are not reliable. In general, over-diagnosis increases with age across all age groups, but rises more sharply for women who are screened in their 70s and 80s.

[3] Shaded strategies are dominated by other strategies; the strategy that dominates may not be on the table.

Table 5b

Summary of Benefits and Harms-Comparison of Different Stopping Ages Using Exemplar Model [1]

		Potential Benefits (vs. no screening)			Potential Harms[2]	
Strategy	**Average Screens/1000**	**% Mortality Reduction**	**Cancer Deaths Averted / 1000**	**Life Years Gained/1000**	**# False positives/1000**	**# of unnecessary biopsies/1000**
Biennial						
B 50–69	8944	15%	5.4	99	780	55
B 50–74	11109	20%	7.5	121	940	66
B 50–79	12347	25%	9.4	130	1020	71
B 50–84	13836	26%	9.6	138	1130	79
Annual						
A 50–69	17759	20% [3]	7.3	132	1350	95
A 50–74	21357	26%	9.5	156	1570	110
A 50–79	24439	30%	11.1	170	1740	122
A 50–84	26913	33%	12.2	178	1880	132

A= Annual

B= Biennial

[1] Results are from Model S. Model S was chosen as an exemplar model to summarize the balance of benefits and harms associated with screening 1000 women under a particular screening strategy..

[2] Over-diagnosis is another significant harm associated with screening. However, given the uncertainty in the knowledge base about DCIS and small invasive tumors, we felt that the absolute estimates are not reliable. In general, over-diagnosis increases with age across all age groups, but rises more sharply for women who are screened in their 70s and 80s.

[3] Shaded strategies are dominated by other strategies; the strategy that dominates may not be on the table (see other tables).

Appendix Table 1

Summary of Model Features

	D	E	G	M	S	W
Includes DCIS	No	Yes	Yes	Yes	No	Yes
Includes ER status	Yes	Yes	Yes	Yes	Yes	Yes
How treatment affects mortality	Hazard reduction	Cure fraction	Hazard reduction	Hazard reduction and cure fraction based on mode of diagnosis *	Hazard reduction	Cure fraction
Calibrated to mortality?	No	No	No	Yes	No	Yes **
Calibrated to incidence?	No	Yes	Yes	Yes	Yes	Yes
Factors affecting screening benefits ***	Stage shift, age shift	Size (larger or smaller than fatal diameter)	Stage shift, age shift	Stage shift, age shift	Stage shift, size within stage, age shift	Effectiveness of treatment by stage and age shifts
Factors affecting treatment benefits (independent of screening)	ER, age, calendar year	ER, age	ER, age	ER, age, calendar year (and improvements in care)	ER, age	ER, age, calendar year (which affect cure probability)

Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G (Georgetown U.), M (M.D. Anderson Cancer Center), S (Stanford U.), W (U. of Wisconsin/Harvard)

ER = estrogen receptor

* If cancer is clinically detected in model M, a hazard reduction is applied to the survival function. If a cancer is screen detected then a cure fraction is applied for cases diagnosed in stages 1 and 2a. If a cancer is screen detected in stages 2b, 3, 4 a similar hazard reduction is applied as for the clinically detected cases. This results in screening benefits due to stage shift and better prognosis for screen vs. clinical cases within early stage disease. The use of a cure fraction for early stage screen detected cancer represents a modification from the previously published model. (reference 7;11)

** Model W is only calibrated to mortality for a subset of the cure fraction parameters after the natural history model was calibrated to incidence.

*** Note that all models use age-specific inputs for sensitivity of mammography screening; sensitivity, in turn, has some small effect on screening benefits.

Appendix Table 2

Summary of Base Case Input Data Sources

Model inputs			Datasets*			
			BCSC	**SEER 9**	**Connecticut Tumor Registry**	**Berkeley Mortality Database**
Secular breast cancer incidence				✓	✓	
Mammography test characteristics			✓			
Other-cause mortality						✓
1975 breast cancer survival				✓		
1975 breast cancer prevalence				✓	✓	

*Abbreviations: NHIS, National Health Interview Survey; SEER POC, SEER Patterns of Care; BCSC, Breast Cancer Surveillance Consortium; SEER 9, Surveillance, Epidemiology and End-Results Nine Registries.

For this analysis we assume that 100% of women are screened and that all women detected with cancer are treated as per current practice guidelines.

Appendix Table 3

The Average Number of Screening Exams per 1000 Woman and the Gain in Years of Life per 1000 Women Screened for each Model by Screening Strategy

Screening Strategies	Average # of screens per 1000women[1]	Years of Life Gained per 1000 Women (vs. no screening) Models[2]					
		D	E	G	M	S	W
Efficient Strategies (not dominated in 5 or 6 of 6 models)							
B 60–69	4263	51	49	61	43	52	39
B 55–69	6890	73	78	91	62	80	64
B 50–69	8947	88	107	111	82	99	84
B 50–74	11066	106	116	128	96	121	95
B 40–79	17241	133	161	164	122	151	136
B 40–84	18708	140	164	167	126	158	140
A 40–79	34078	170	224	188	123	202	198
A 40–84	36550	177	227	192	128	210	202
Borderline Strategies (dominated in 2–4 of 6 models)							
B 45–69	11694	102	129	136	99	116	109
B 50–79	12366	114	122	136	103	130	99
B 50–84	13837	121	124	139	108	138	103
B 40–69	13831	108	147	140	101	120	121
A 45–69	22546	131	179	152	103	152	155
A 50–79	24419	145	166	154	112	170	142
A 50–84	26905	152	169	157	116	178	146
A 40–69	27428	142	206	162	103	164	180
Dominated Strategies (dominated in all 6 models)							
A 60–69	8438	65	69	71	53	69	56
A 55–69	13009	91	107	100	68	102	90
A 50–69	17733	117	148	128	91	132	123
A 50–74	21330	134	160	144	104	156	135

A=Annual

B=Biennial

[1]Average number of mammograms across models. Not all possible mammograms in the age interval are obtained in strategies that continue to the oldest age groups since many women die as the result of other causes before screening would occur.

[2]Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G (Georgetown U.), M (M.D. Anderson Cancer Center), S (Stanford U.), W (U. of Wisconsin/Harvard)

[3]Shaded areas in the table show strategies that are dominated within a specific model; a strategy is classified as dominated if there is another strategy (from either the Efficient, Borderline or Dominated categories) that results in an equal or higher years of life gained with fewer average screening exams.