

Using Collaborative Simulation Modeling to Develop a Web-Based Tool to Support Policy-Level Decision Making About Breast Cancer Screening Initiation Age

Elizabeth S. Burnside, MD, MPH, MS, Sandra J. Lee, ScD, Carrie Bennette, PhD, Aimee M. Near, MPH, Oguzhan Alagoz, PhD, Hui Huang, MS, Jeroen J. van den Broek, MS, Joo Yeon Kim, MS, Mehmet A. Ergun, MSc, Nicolien T. van Ravesteyn, PhD, Natasha K. Stout, PhD, Harry J. de Koning, MD, PhD, Jeanne S. Mandelblatt, MD, MPH

Background: There are no publicly available tools designed specifically to assist policy makers to make informed decisions about the optimal ages of breast cancer screening initiation for different populations of US women. **Objective:** To use three established simulation models to develop a web-based tool called Mammo OUTPuT. **Methods:** The simulation models use the 1970 US birth cohort and common parameters for incidence, digital screening performance, and treatment effects. Outcomes include breast cancers diagnosed, breast cancer deaths averted, breast cancer mortality reduction, false-positive mammograms, benign biopsies, and overdiagnosis. The Mammo OUTPuT tool displays these outcomes for combinations of age at screening initiation (every year from 40 to 49), annual versus biennial interval, lifetime versus 10-year horizon, and breast density, compared to waiting to start biennial screening at age 50 and continuing to 74. The tool was piloted by decision makers ($n = 16$) who completed surveys. **Results:** The tool demonstrates

that benefits in the 40s increase linearly with earlier initiation age, without a specific threshold age. Likewise, the harms of screening increase monotonically with earlier ages of initiation in the 40s. The tool also shows users how the balance of benefits and harms varies with breast density. Surveys revealed that 100% of users (16/16) liked the appearance of the site; 94% (15/16) found the tool helpful; and 94% (15/16) would recommend the tool to a colleague. **Conclusions:** This tool synthesizes a representative subset of the most current CISNET (Cancer Intervention and Surveillance Modeling Network) simulation model outcomes to provide policy makers with quantitative data on the benefits and harms of screening women in the 40s. Ultimate decisions will depend on program goals, the population served, and informed judgments about the weight of benefits and harms. **Key words:** simulation modeling; breast cancer screening; health care policy; decision making; mammography. (*MDM Policy & Practice* 2017;2:1–10)

The age of screening mammography initiation remains the focus of substantial policy debate. The US Preventive Services Task Force recommends that screening before age 50 be individualized based on a woman's personal risk of breast cancer and preferences.^{1,2} The American Cancer Society recently published guidelines that recommend annual screening

from ages 45 to 54.³ Specialty organizations such as the American College of Radiology and the American College of Obstetrics and Gynecology recommend annual screening starting at age 40.^{4,5} These variations are largely due to differences in the weights placed on evidence from clinical trials, observational studies, and the balance of benefits and harms for different groups of women.^{3,6,7}

This inconsistency in recommended breast cancer screening initiation age leaves decision makers such as health insurers, clinicians, state and local health departments, and other professional societies with

© The Author(s) 2017

Reprints and permission:

<http://www.sagepub.com/journalsPermissions.nav>

DOI: 10.1177/2381468317717982



This Creative Commons Non Commercial CC-BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

considerable uncertainty in making pragmatic decisions about how to implement screening in their organization or target population. Furthermore, while numerous decision aids exist to support individual decisions in breast cancer^{8–10} and lung cancer^{11,12} screening, there are fewer tools designed for policy-level decisions. Web-based tools are available for policy makers to evaluate the expected population-level impacts from alternative colorectal cancer screening and treatment options,¹³ evidence

supporting public health initiatives,¹⁴ or methods for health care financing in general,¹⁵ but none evaluate the population-level impact of different breast cancer screening strategies with particular attention to initiation age.

To fill this gap, we developed an interactive, web-based tool called the Mammography Outcomes Policy Tool (Mammo OUTPuT) that integrates policy-level parameters and outcomes from three well-established Cancer Intervention and Surveillance Modeling Network (CISNET) simulation models. Previously, these models only simulated outcomes for two ages (40 and 45) within this age range. The Mammo OUTPuT tool allows policy makers to vary population characteristics (age and breast density) and screening intervals (annual and biennial) to quantitate the trade-offs inherent in practice decisions for their specific populations of interest. This tool is intended to ultimately provide data that can be used by diverse policy-making audiences in their decisions about breast cancer screening guidelines.

METHODS

Model Overview

We constructed the Mammo OUTPuT tool using three simulation models developed independently within CISNET. This research is institutional review board exempt since all data were de-identified. The simulation models include the following: model D (Dana-Farber Cancer Institute, Boston, Massachusetts), model E (Erasmus Medical Center, Rotterdam, the Netherlands), and model W (University of Wisconsin, Madison, Wisconsin, and Harvard Medical School, Boston, Massachusetts).^{16–18} Full details of the models are available at <https://resources.cisnet.cancer.gov/registry>.¹⁹

Briefly, the models begin with estimates of breast cancer incidence²⁰ and ER/HER2-specific survival trends *without* screening or adjuvant treatment and then overlay data on screening and molecular subtype-specific adjuvant treatment to generate observed US population incidence and mortality trends.^{21–25} Breast cancers have a distribution of preclinical screen-detectable periods (sojourn time) and clinical detection points. Screen detection of cancer during the preclinical screen-detectable period can result in the identification (and treatment) of earlier-stage or smaller tumors than might occur via clinical detection, with a corresponding improvement in breast cancer mortality and

Received 8 February 2017 From the Breast Cancer Working Group of the Cancer Intervention and Surveillance Modeling Network (CISNET). Three independent modeling teams did this work from Dana-Farber Cancer Institute, Harvard Medical School (Pl: Lee); Erasmus Medical Center (Pl: de Koning); Harvard Medical School, University of Wisconsin (Pl: Trentham-Dietz/Stout/Alagoz). Elizabeth Burnside, Jeanne Mandelblatt, Sandra Lee, and Aimee Near were the writing and coordinating committee for the project. From the Department of Radiology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin (ESB); Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts (SJL, HH); Group Health Research Institute, Seattle, Washington (CB); Department of Oncology, Georgetown University Medical Center, and Cancer Prevention and Control Program, Georgetown-Lombardi Comprehensive Cancer Center, Washington, DC (JSM, AMN); Carbone Cancer Center, University of Wisconsin, Madison, Wisconsin (OA, MAE); Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, Netherlands (HJK, NTR, JJB); Department of Health Policy and Management, Graduate School of Public Health, University of Pittsburgh, Pittsburgh (JYK); Department of Population Medicine, Harvard Medical School, and Harvard Pilgrim Health Care Institute, Boston, Massachusetts (NKS). Financial support for this study was provided by the National Institutes of Health under National Cancer Institute Grant U01 CA152958 and a contract to Cornerstone supported under NCI Grant U01 CA152958. This work was also supported in part by National Institutes of Health grants R01CA165229 and K24CA194251 as well as the University of Wisconsin Carbone Cancer Support Grant P30CA014520 (EB), by grants and contracts that support the Breast Cancer Surveillance Consortium (P01CA154292, U54CA163303, HHSN261201100031C), and by the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS), Grant UL1TR000427. The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Revision accepted for publication 25 April 2017.

The online appendixes for this article are available on the *Medical Decision Making Policy & Practice* Web site at <http://journals.sagepub.com/home/mpp>.

Address correspondence to Elizabeth Burnside, MD, MPH, MS, Department of Radiology, University of Wisconsin School of Medicine and Public Health, 600 Highland Ave, Madison, WI 53792, USA; telephone: (608) 263-8340; e-mail: eburnside@uwhealth.org.

life-years gained. Digital mammography performance characteristics are based on age (grouped as 40–49, 50–64, 65+), first versus subsequent screen, time since last mammogram (annual, biennial), and breast density. Women can die of breast cancer or other causes.

A summary of the assumptions about risk related to breast density, natural history, and the data from which they were obtained are covered in full in previously published sources.^{19,26} Briefly, the age-specific prevalence of breast density was determined from Breast Cancer Surveillance Consortium (BCSC) data from 1994 to 2010. Density, in turn, affected age-specific risk of development of breast cancer based on risk ratios from the BCSC data. Third, density also affected the sensitivity of digital mammography based on data from the BCSC.

The natural history of DCIS (ductal carcinoma in situ) is an unobservable phenomenon. Each model makes slightly different assumptions about DCIS. In general, the models all assume that a certain proportion of DCIS are not destined to progress; the remainder will progress to invasive cancer. The rates of each type of DCIS were determined by calibration using combinations of several variables: observed DCIS incidence rates over time, screening test performance for DCIS, assumptions about tumor growth rates, and other model-specific parameters. All DCIS, progressive and nonprogressive, can be screen detected. Screen detection of a DCIS that was either never destined to progress or would have never been detected in the absence of screening due to death from other causes are considered overdiagnosis. Likewise, invasive cases that would not ever have been detected in the absence of screening due to death from other causes, or in the case of Model W, due to nonprogression of a small percentage, would also be overdiagnosis.

We simulate a cohort of women born in 1970 and follow them from age 25 (since breast cancer is rare before this age [0.08% of cases]) until death or age 100. We select the 1970 cohort since this is the group that was age 40 in 2010, making it ideal to assess outcomes related to current screening initiation decisions.

Simulation Model Input Parameters

The three models begin with a common set of age-specific variables for breast cancer incidence, digital mammography performance characteristics, ER/HER2-specific treatment effects, and non-breast

cancer competing causes of death.¹⁹ In addition, on the basis of their specific model structure each group includes model-specific inputs (or intermediate outputs) to represent preclinical detectable times, lead-time, as well as age- and ER/HER2-specific stage distribution in screen- versus non-screen-detected women.^{20–28} The models assume 100% adherence to screening and the most effective treatment to quantify the efficacy of screening strategies. Results are tabulated for each model by calculating the within-model differences between each screening strategy and no screening. All model input parameters are available at <https://resources.cisnet.cancer.gov/registry>.¹⁹

The models quantify outcomes for four breast density subgroups as defined by the American College of Radiology Breast Imaging Reporting and Data Systems (BI-RADS): a = almost entirely fatty; b = scattered fibroglandular; c = heterogeneously dense; d = extremely dense as well as all breast density categories combined (a group that we will heretofore refer to as the “combined average density”). Breast density is assigned at age 40 years and can decrease one level or remain the same at age 50 and again at age 65 years using the age-specific density prevalence rates from the BCSC.¹⁹ Density-specific digital mammography sensitivity and specificity based on age, screening round, and screening interval are estimated from the BCSC data.¹⁹ Screening interval uses standard BCSC definitions: annual includes data from screens occurring within 9 to 18 months of the prior screen and biennial includes data on screens within 19 to 30 months. Density also modifies age-specific risk of developing breast cancer. The models incorporate this risk for age groups 40 to 49, 50 to 64, and 65+ years using the combined average density-related risk in each age group as the referent group.¹⁹ The simulation models enable quantification of outcomes subsequently used in the Mammo OUTPuT tool (see Online Table 1).

Benefits

Outputs related to screening benefits included in the Mammo OUTPuT include breast cancers diagnosed (total, invasive, and DCIS), breast cancer deaths averted, percent breast cancer mortality reduction, and life-years gained.^{19,29–31} Benefits (and harms) are accumulated from age at screening program initiation through age 99 years to capture the lifetime impact of screening strategies.

Harms

Harms include false-positive mammograms, benign biopsies, and overdiagnosis.¹⁹ A false-positive mammogram is defined as a mammogram read as abnormal and needing further work-up in a woman without cancer. A benign biopsy is defined as a biopsy recommendation for a woman with false-positive screening results. Overdiagnosis is defined as a cancer that would not have been clinically detected in the absence of screening (because of lack of progressive potential or death from competing mortality). Percent overdiagnosis is estimated using the total number of breast cancer diagnoses for a specified horizon as a denominator.

The Mammography Outcomes Policy Tool (Mammo OUTPuT)

Our web-based, policy-level tool (see screen capture example in Online Figure 1) displays a series of interactive figures to communicate the results of the simulation models for initiating screening at each individual year of age between 40 and 49.³² The tool presents the results of over 100 different scenarios for screening initiation by varying the outcome of interest, screening interval, horizon, and breast density. Outputs are further varied by screening initiation age (40, 41, . . . , 48, 49), allowing the tool to support visualization of more than 2,000 combinations.

Outcomes from each simulated scenario are compared to those expected without any screening to generate the results for a given analysis. The results for strategies for each starting age in the 40s is then compared to results for the same cohort if screening had not started until age 50 (and continue to age 74) to estimate the impact of earlier initiation. Results for the models are depicted as a median.

The data are then displayed graphically. The primary graphic, a simple bar chart that illustrates the selected outputs, is used as a familiar and easy to understand format to view the results. To facilitate comparisons across different starting ages for screening mammography in the 40s, screening initiation age is shown on the x-axis and results of the selected outcome shown on the y-axis. Results for starting at age (40 + n , where $n = 0$ to 9) are compared to those expected if women with those characteristics had waited until age 50 to start screening biennially.

Usability Testing

A preliminary version of the tool (shown in the video available here: <https://www.hipxchange.org/MammoOUTPuTVideo>; username: mammooutput, password: review) was pilot tested in a convenience sample of decision makers, clinicians, and breast cancer researchers. Users recorded the time they spent investigating the tool and completed a survey (Appendix 1).

RESULTS

We successfully designed, constructed, revised (according to usability testing), and posted the final Mammo OUTPuT tool to a publically available website: <https://www.hipxchange.org/MammoOUTPuT>. The UW-Madison Health Innovation Program (HIP) supports a web portal called the HIPxChange, which provides the infrastructure to disseminate research results. The link provides users with a username and password and then allows access to the Mammo OUTPuT tool as well as information on how to use the tool and how to interpret the tool results.

In the results section we present a summary of pilot usability testing that we performed prior to posting the final tool online. In addition, in the results section, we summarize components of the tool not previously published, which are uniquely communicated via the interactive and visual nature of the tool.

Pilot Usability Testing

A total of 16 decision makers, clinicians, and breast cancer researchers pilot tested the preliminary tool. They spent a mean of 44 minutes (range 25–120 minutes) exploring the tool. All respondents liked the appearance of the site; 88% (14/16) stated that the website was either “very easy” or “extremely easy” to navigate; 94% (15/16) found the website helpful for their practice; and 94% (15/16) would recommend the tool to a colleague. Users felt that deaths avoided, mortality reduction, life years gained, false-positive mammograms, benign biopsies, and overall overdiagnosis numbers were the most important outcomes, while total number of breast cancer diagnoses (incidence) and overdiagnosis presented separately as invasive and DCIS were viewed as less important. Other outcomes that users

requested were quality-adjusted life years as well as “life years gained per exam.” Since these outcomes were not directly available from the models, we could not make this change. A few users initially found the instructions difficult to comprehend prompting comments such as “instructions are too wordy” and “instructions are lengthy and hard to read. I had to reread sentences a few times to grasp concepts.” We have rewritten the instructions in the current tool to address these concerns. Several users found the graphics challenging to comprehend, articulating that “bars depicting the benefit of screening biennially starting at age 50” were confusing. Specifically, the labeling on the x-axis implies these outcomes occur during the 40 to 49 age range rather than in later years. We used this input to reconfigure the graphics. Finally, several users expressed a desire to view multiple scenarios side-by-side for easier comparison, which we now provide.

New Concepts Presented by Mammo OUTPuT

For all combinations of age, density, and screening interval, the tool enables the user to visualize that there is a monotonic trend in breast cancer outcomes across ages in the 40s without a clear cut-point (Online Figure 1a, left graphic). Likewise, the harms are inversely related to age of screening initiation in a similar monotonic pattern (Online Figure 1a, right graphic). The tool allows comparison of specific ages, in order to drill down on policies of interest, for example, comparing initiation ages of 40 and 45 as compared to waiting to start screening biennially from 50 to 74 (Online Figure 1b).

The tool visually demonstrates subtle details underlying summarized outcomes; nuances that might not be fully appreciated if the outcomes were only viewed in tabular form (Online Table 2). For example, when viewing the number of cancers (invasive + in situ) diagnosed per 1,000 women, only 2.9 additional cancers are diagnosed when screening annually from age 40 compared to waiting to start screening biennially from 50 to 74. However, beginning screening earlier than age 50 can shift some detected invasive cancers to DCIS, creating a relative deficit of subsequent invasive cases. This stage shift effect will only avert breast cancer deaths to the extent that DCIS progresses to invasion. The DCIS cases not destined to progress will result in cases of overdiagnoses/overtreatment. Mammo OUTPuT can demonstrate this shift for all or selected initiation ages between 40 and 49 (Online Figure 2).

Mammo OUTPuT also provides new insights into the outcome differences depending on breast density. When comparing outcomes of screening in the 40s for all women (combined average density) to women with extremely dense breasts, interesting trends emerge when viewing these results in tabular form (Online Table 3), but these patterns are powerfully illustrated in graphical form (Online Figure 3). Specifically, screening for women with extremely dense breasts results in more accrued benefits, while accrued harms stay virtually the same. An example of these trends in terms of benefits are summarized visually in Mammo OUTPuT by comparing life years gained in all women (Online Figure 3a) and those with extremely dense breasts (Online Figure 3b). An example of these trends in terms of harms are summarized visually in Mammo OUTPuT by comparing overdiagnosis in these same density scenarios (Online Figure 4a and b).

DISCUSSION

The Mammo OUTPuT tool is the first web-based decision tool that enables policy decision makers to visualize and quantify the outcomes of mammography screening in the 40s based on specific initiation age, breast density, and screening interval. This is the first time that outcomes are available for every year within this age range. The visualization of outcomes provided by the tool illustrates, as suspected based on prior results, that there are no cut-points of age where choices are obvious in terms of benefits or harms. Rather, the choice is dependent on program goals, the population served, and the value placed on the relative weight of benefits and harms of mammography screening. Pilot testing of the tool demonstrated the preliminary acceptability, usability, and utility to a range of decision makers.

While there is not complete consensus on breast cancer screening guidelines,^{3,7} there is broad agreement that screening women in the 40s has some benefit in terms of breast cancer mortality reductions and breast cancer deaths averted.^{33,34} However, the overall magnitude of benefit observed in clinical trials and observational studies is less than in older age groups,^{33–35} making screening initiation decisions more complex and value-based.³⁶ The goal of this tool is to provide diverse policy decision makers with data to translate simulation results in a timely, relevant, and easily accessible manner.³⁷ Mammo OUTPuT contributes a unique, interactive method to understand screening

outcomes for every year between 40 and 49 providing policy makers with perspectives not previously available. Since specific outcomes (and the balance of benefits and harms) vary by combinations of factors, we present hypothetical scenarios to demonstrate how various decision makers might use this tool to inform their decisions.

Breast Cancer Policy Decisions for an Integrated Health Plan

Directors of integrated health plans must make decisions about provision of services for their covered population weighing population characteristics, resources, and competing health needs. In this situation, the Mammo OUTPuT tool could be used by a director of an integrated health plan responsible for a rural population with a younger than average age distribution. As shown above (Online Figure 1) for the average US population, adopting a breast cancer screening initiation at age 40 would avert the most deaths, but also induce the most potential false-positives, perhaps, requiring referral into a more urban area for follow-up diagnostic procedures.

The director might be concerned that compared to the average US population, his/her covered population is young and includes a large number of women with dense breasts who have an increased risk of disease. The director could examine the outcomes for women with extremely dense breasts (Online Figures 3 and 4 and Table 3) and estimate the outcomes over a lifetime horizon. The tool shows that annual screening in the 40 to 49 age group with extremely dense breast tissue avoids more cancers and deaths and incurs fewer false-positives, biopsies, and overdiagnosis as compared with the density distribution of all women (combined average density). Thus, for women with extremely dense breast tissue, this policy maker may elect annual mammography starting at age 40, deciding that the greater number of deaths averted, but lower rates of benign biopsies make this a reasonable strategy for this specific group, while choosing another strategy for women who do not have extremely dense breast tissue.

Screening Decision Making by Consumer Advocacy Organizations

There are consumer advocacy groups interested in the specific needs of women based on their breast

density,³⁸ or those interested in ensuring that women avoid overdiagnosis and unnecessary treatment.³⁹ A director of an advocacy organization whose primary mission was to avoid all possible breast cancer deaths could use the tool to select the screening strategy that maximized mortality reduction and life years saved. For this goal, the results from the tool suggest promotion of annual screening initiation at age 40 (Online Figure 3). However, for an organization whose priority was to avoid unnecessary treatment, the tool provides data to determine the balance of breast cancer deaths averted relative to added cases of over-diagnosis and over-treatment (Online Figure 4).

Screening Decisions by Public Program Directors

A decision maker may have a fixed budget to provide services such as is the case in local departments of health or the Centers for Disease Control and Prevention's National Breast and Cervical Cancer Early Detection Program.⁴⁰ Often the population targeted by these programs is younger and more underserved than the average US female population eligible for screening. In this instance, the decision maker may want to know about the proportion of benefits captured for an average population from ages 40 to 49 if screening is provided on an annual versus a biennial basis. As shown in Online Table 3, the tool demonstrates that biennial screening would preserve 87% of the benefits in terms of life years gained (42.7 v. 37.1) in women with combined average density. Therefore, the decision maker might decide that they could implement a biennial program allowing coverage of twice as many women as could be served under an annual program with only a small trade-off in terms of loss of potential life years gained.

Setting Professional Guidelines

Another group of decision makers that might be users of this tool include those are tasked with developing guidelines for their professional subspecialty group. Sometimes professional groups will adopt prevailing guidelines, like those published by the American Cancer Society³ or the US Preventive Services Task Force.⁷ However, the subspecialty guideline decision maker may feel that their organization serves women that differ from those in the general population. For instance, breast surgeons often care for women with dense breasts referred for

evaluation for biopsy that continue to return for follow-up. A population with a higher breast density distribution as compared to the expected breast density distribution would have a higher risk for breast cancer.⁴¹ Therefore, a breast surgery decision maker might recommend annual screening beginning at age 40 for women seen by their specialty.

A growing number of organizations in the public and private sectors now rely on simulation modeling to better understand the health and economic consequences of alternative policy decisions.⁴² However, few use collaborative modeling or make their results available to policy makers in an accessible, web-based format that allows manipulation by the user—as provided in the Mammo OUTPuT tool enabled by the CISNET breast consortium. The Colorectal Cancer Mortality Projections website is another notable example of an interactive tool that projects collaborative simulation modeling results. However, this tool provides insight into a single outcome, colorectal cancer mortality, which depends on interventions including risk factor reduction, early detection, and/or increased access to optimal treatment.¹³ In contrast, the Mammo OUTPuT tool helps decision makers consider how different early detection strategies will affect a broad range of outcomes, including breast cancer deaths, the number of biopsies and false-positive screens, and the number of overdiagnosed cases, among others. The choice of preferred outcome will vary based on a decision maker's mandate and context. Thus, Mammo OUTPuT provides greater flexibility for decision makers to consider the outcomes most relevant to their population and mission.

Mammo OUTPuT is a policy-level decision tool, which differs in scope and objective from patient decision aids that are now commonly used to help women make individualized decisions regarding breast cancer screening. In contrast to a patient decision aid, our tool takes a population-level perspective by illustrating the benefits and harms over a large relevant patient population rather than for a single patient. Though benefits and harms may overlap with those considered important by patients and therefore included in patient decision aids, they differ in how the information is presented. For example, both our tool and several patient decision aids^{8–10} present quantitative information about false-positive mammography results. However, our tool focuses on illustrating the total number of false-positives in a *cohort* of women over time. An individual patient decision aid focuses on the likelihood an *individual* will experience a false-positive

result and on providing patient-centered contextual information to help women understand these outcomes (e.g., information about how the extra tests and waiting time associated with a false-positive result can cause anxiety in some women). Another reason that the tool is only appropriate for a population is related to the data underpinning the model. There are substantial correlations between mammograms performed on the same women, correlations that were not available in the data on which the model is built. Thus, the tool graphics truly represent outcomes for a cohort rather than outcomes for any given individual.

The preliminary usability data suggest that this tool has potential to provide interactive breast cancer screening outcomes from simulation models to users. The impressions from surveys of our small convenience sample are encouraging, indicating that users like the tool, would use the tool, and would recommend the tool to others. However, the small sample size and limited number of policy makers included limits the strength of conclusions that we can draw. Additional study of the information effectively conveyed by the tool to users, with a larger and more inclusive survey and/or perhaps in-depth interviews, would add to our understanding. For example, will the tool change a user's mind about breast cancer screening initiation age or will this information help them make policy decisions? Further work to address the feedback regarding user instructions and alternative methods to compare graphical depiction of the outcomes data across scenarios will be important areas for future enhancements.

We have attempted to design our tool to focus on outcomes considered most important to the diverse audience of policy makers, health insurers, and state and local health departments; however, there is little literature on which to base these judgments. The set of outcomes included in the tool was informed by prior interactions between the CISNET team and those who set guidelines and policy recommendations, such as the US Preventive Services Task Force and several large health insurers, but we acknowledge that this list may not be exhaustive of all outcomes considered important by all potential users (e.g., cost, health-related quality of life, quality-adjusted life years, or patient preferences). Understanding the tool characteristics most valued by the targeted audience is considered important future work.

Overall, the Mammo OUTPuT web-based tool uses well-established models and modern data on breast cancer to support evidence-based policy

decisions and clinical practice guidelines by facilitating the direct comparison of key outcomes under alternative mammography screening strategies. However, there are several caveats that should be considered in evaluating the tool. First, this tool is not intended for use in individual clinical decision making as discussed previously. There are other web-based decision aids that address some aspects of screening decisions for young women.^{8–10} Second, the tool was designed to address screening initiation decisions in the context of the US system perspective only. We do not provide data on different intervals or strategies for women 50 and older nor do we consider conventions that are in place (e.g., triennial screening) in other national screening programs. Next, the tool does not include data on risk factors for breast cancer other than increased breast density. Many are now suggesting that risk-tailored strategies be considered in future guidelines as evidence evolves in this area.^{43,44} Furthermore, this tool does not incorporate a key component of breast cancer screening guidelines, that screening decisions be individualized to reflect a woman's values and preferences.^{3,7} Additionally, the tool does not consider the costs of screening and downstream events, so it cannot be used to directly evaluate the budget impact of different policy decisions. This will be an important area for future expansion. Next, the tool assumes 100% adherence to screening, prompt evaluation of abnormal results, and full use of optimal treatment to evaluate program efficacy. Decision makers using the tool should be cognizant of the fact that actual benefits (or harms) may not match projected results. For example, benefits may fall short of the projected results since adherence to both screening and treatment is not perfect. In future work, we will be adding options to model adherence patterns. In addition, the tool provides the median estimate from the three models for ease of visualization. In future refinements, the tabular data will include the median and the range of results across the models. These models have generated very consistent outcomes in the past^{19,24,25}; therefore, the range data should not affect conclusions about starting ages based on the median alone. In future tool expansions, it will also be important to include other potential outcomes of interest to policy makers like quality-adjusted life years. While the models depict outcomes for 1-year age groupings, some input parameter data are only available collapsed across 5- or 10-year intervals, decreasing the differences across ages. Finally, while the models underlying the tool

are well established and accurately reproduce US incidence and mortality trends and results of screening trials in younger women,¹⁹ the models make some assumptions about unobservable events in the natural history of breast cancer (e.g., the proportion of DCIS cases that are not destined to progress). The consistency of within and across model analyses results for this tool and in other model-based analyses using the same input parameters¹⁹ should provide greater confidence in results than tools based on one model.

Overall, the Mammo OUTPuT tool has several important strengths including collaboration of three independent modeling groups using modern screening data including breast density,¹⁹ interactive results, and outcomes previously used to influence policy. This tool should enable users to visualize the trade-offs in terms of the benefits and harms of screening mammography and contribute to more informed policy decisions.

ACKNOWLEDGMENTS

The authors acknowledge the work of Cornerstone Systems Northwest, Inc., in developing the web interface and UW Madison HIPxChange for assisting in creation of the toolkit and providing online access to the tool. In addition, the authors acknowledge the Breast Cancer Surveillance Consortium (BCSC) investigators as well as state public health departments and cancer registries throughout the United States that provide cancer and vital status data to the BCSC. For a full description of BCSC investigators and these sources, please see: <http://breast-screening.cancer.gov/work/acknowledgement.html>.

REFERENCES

1. US Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2009;151(10):716–26, W-236.
2. Siu AL; US Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2016;164(4):279–96.
3. Oeffinger KC, Fontham EH, Etzioni R, et al. Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. *JAMA*. 2015;314(15):1599–614.
4. American College of Obstetricians-Gynecologists. Practice bulletin no. 122: breast cancer screening. *Obstet Gynecol*. 2011; 118(2 Pt 1):372–82.
5. Lee CH, Dershaw DD, Kopans D, et al. Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *J Am Coll Radiol*. 2010;7(1):18–27.

6. Kerlikowske K. Progress toward consensus on breast cancer screening guidelines and reducing screening harms. *JAMA Intern Med.* 2015;175(12):1970–1.
7. US Preventive Services Task Force. Draft recommendation statement breast cancer 2015. Available from: www.uspreventiveservicestaskforce.org/Home/GetFile/6/250/bcdraftrecbulletin/pdf
8. Breast Cancer Surveillance Consortium. Breast Cancer Surveillance Consortium risk calculator. Available from: <https://tools.bccsc-scc.org/BC5yearRisk/calculator.htm>
9. Weill Medical College. Breast screening decisions. Available from <http://breastscreeningdecisions.com/#/>
10. Scariati P, Nelson L, Watson L, Bedrick S, Eden KB. Impact of a decision aid on reducing uncertainty: pilot study of women in their 40s and screening mammography. *BMC Med Inform Decis Mak.* 2015;15:89.
11. Lau YK, Caverly TJ, Cherng ST, et al. Development and validation of a personalized, web-based decision aid for lung cancer screening using mixed methods: a study protocol. *JMIR Res Protoc.* 2014;3(4):e78.
12. University of Michigan. Lung cancer online risk assessment and screening tool. Available from: <http://www.shouldiscreen.com/how-is-screening-done>
13. National Cancer Institute. Colon cancer online tool. Available from: http://cisnet.cancer.gov/projections/colorectal/compare_scenarios.php
14. Ainsworth JD, Carruthers E, Couch P, et al. IMPACT: a generic tool for modelling and simulating public health policy. *Methods Inf Med.* 2011;50(5):454–63.
15. Skordis-Worrall J, Pulkki-Brannstrom AM, Utley M, et al. Development and formative evaluation of a visual e-tool to help decision makers navigate the evidence around health financing. *JMIR Res Protoc.* 2012;1(2):e25.
16. 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.
17. Lee S, Zelen M. A stochastic model for predicting the mortality of breast cancer. *J Natl Cancer Inst Monogr.* 2006;(36):79–86.
18. 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.
19. Mandelblatt JS, Cronin KA, De Koning HJ, Miglioretti DL, Schechter CB, Stout NK. Collaborative modeling of U.S. breast cancer screening strategies. Available from: <http://www.uspreventiveservicestaskforce.org/Page/Document/modeling-report-colaborative-modeling-of-us-breast-cancer-1/breast-cancer-screening1>
20. Gangnon RE, Sprague BL, Stout NK, 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;24(6):905–12.
21. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med.* 2005;353(17):1784–92.
22. Chang Y, Schechter CB, van Ravesteyn NT, et al. Collaborative modeling of the impact of obesity on race-specific breast cancer incidence and mortality. *Breast Cancer Res Treat.* 2012;136(3):823–35.
23. Munoz D, Near AM, van Ravesteyn NT, et al. Effects of screening and systemic adjuvant therapy on ER-specific US breast cancer mortality. *J Natl Cancer Inst.* 2014;106(11). doi: 10.1093/jnci/dju289.
24. van Ravesteyn NT, Miglioretti DL, Stout NK, 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.
25. Mandelblatt JS, Cronin KA, Bailey S, 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.
26. Mandelblatt JS, Stout NK, Schechter CB, et al. Collaborative modeling of the benefits and harms associated with different U.S. breast cancer screening strategies. *Ann Intern Med.* 2016;164(4):215–25.
27. 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.
28. Holford TR, Cronin KA, Mariotto AB, Feuer EJ. Changing patterns in breast cancer incidence trends. *J Natl Cancer Inst Monogr.* 2006;(36):19–25.
29. 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(4):538–44.
30. 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.
31. 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.
32. Speyer P, Pagels B, Park N. A white paper published by the Institute for Health Metrics and Evaluation. Available from: http://communicatingdata.org/wp-content/uploads/2014/06/Data_Impact_digital.pdf
33. Moss SM, Cuckle H, Evans A, et al. 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.
34. Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. *Lancet Oncol.* 2015;16(9):1123–32.
35. van der Waal D, Broeders MJ, Verbeek AL, Duffy SW, Moss SM. Case-control studies on the effectiveness of breast cancer screening: insights from the UK Age Trial. *Epidemiology.* 2015;26(4):590–6.
36. Siu AL, Bibbins-Domingo K, Grossman D. Evidence-based clinical prevention in the era of the patient protection and Affordable Care Act: the role of the US Preventive Services Task Force. *JAMA.* 2015;314(19):2021–2.

37. Rutter CM, Zaslavsky AM, Feuer EJ. Dynamic microsimulation models for health outcomes: a review. *Med Decis Making*. 2011;31(1):10–8.
38. Capello N. Are you dense? Available from: <http://www.areyoudense.org/>
39. Visco F. New studies for an old story: mammography screening isn't saving lives. *Huffington Post*. Available from: http://www.huffingtonpost.com/fran-visco/new-studies-for-an-old-story-mammography-screening-isnt-saving-lives_b_8033770.html
40. American Cancer Society. Breast cancer early detection and diagnosis. Available from: <http://www.cancer.org/acs/groups/cid/documents/webcontent/003165-pdf.pdf>
41. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med*. 2008;148(5):337–47.
42. National Academies of Sciences. *How Modeling Can Inform Strategies to Improve Population Health: Workshop Summary*. Washington, DC: National Academies Press; 2015. Available from: <http://iom.nationalacademies.org/Reports/2015/How-Can-Modeling-Inform-Strategies-to-Inform-Population-Health.aspx>
43. Onega T, Beaber EF, Sprague BL, et al. Breast cancer screening in an era of personalized regimens: a conceptual model and National Cancer Institute initiative for risk-based and preference-based approaches at a population level. *Cancer*. 2014;120(19):2955–64.
44. Kerlikowske K, O'Kane ME, Esserman LJ. Fifty years of age-based screening: time for a new risk-based screening approach. *Evid Based Med*. 2014;19(5):183.