

# Comparing CISNET Breast Cancer Incidence and Mortality Predictions to Observed Clinical Trial Results of Mammography Screening from Ages 40 to 49

Jeroen J. van den Broek, Nicolien T. van Ravesteyn, Jeanne S. Mandelblatt, Hui Huang, Mehmet Ali Ergun, Elizabeth S. Burnside, Cong Xu, Yisheng Li, Oguzhan Alagoz, Sandra J. Lee, Natasha K. Stout, Juhee Song, Amy Trentham-Dietz, Sylvia K. Plevritis, Sue M. Moss, and Harry J. de Koning

## Abstract

**Background.** The UK Age trial compared annual mammography screening of women ages 40 to 49 years with no screening and found a statistically significant breast cancer mortality reduction at the 10-year follow-up but not at the 17-year follow-up. The objective of this study was to compare the observed Age trial results with the Cancer Intervention and Surveillance Modeling Network (CISNET) breast cancer model predicted results. **Methods.** Five established CISNET breast cancer models used data on population demographics, screening attendance, and mammography performance from the Age trial together with extant natural history parameters to project breast cancer incidence and mortality in the control and intervention arm of the trial. **Results.** The models closely reproduced the effect of annual screening from ages 40 to 49 years on breast cancer incidence. Restricted to breast cancer deaths originating from cancers diagnosed during the intervention phase, the models estimated an average 15% (range across models, 13% to 17%) breast cancer mortality reduction at the 10-year follow-up compared with 25% (95% CI, 3% to 42%) observed in the trial. At the 17-year follow-up, the models predicted 13% (range, 10% to 17%) reduction in breast cancer mortality compared with the non-significant 12% (95% CI, -4% to 26%) in the trial. **Conclusions.** The models underestimated the effect of screening on breast cancer mortality at the 10-year follow-up. Overall, the models captured the observed long-term effect of screening from age 40 to 49 years on breast cancer incidence and mortality in the UK Age trial, suggesting that the model structures, input parameters, and assumptions about breast cancer natural history are reasonable for estimating the impact of screening on mortality in this age group.

## Keywords

breast cancer models, CISNET, external validation, mammography trial simulation

Date received: December 12, 2016; accepted: June 7, 2017

The breast cancer models of the Cancer Intervention and Surveillance Modeling Network (CISNET) synthesize data on breast cancer epidemiology, population demographics, screening accuracy, and treatment effectiveness to simulate the impact of screening and adjuvant treatment interventions on breast cancer incidence and mortality. Prior comparative modeling studies, i.e., cross-validations,<sup>1</sup> by the CISNET models have illustrated the

ability of the models to reproduce the trends in breast cancer incidence and mortality in the United States.<sup>2–4</sup> The models generated similar rankings of the effects of

## Corresponding Author

Jeroen van den Broek, Department of Public Health, Erasmus MC University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands; telephone: +31 107 043 422; fax: +31 107 038 475. (j.j.vandenbroek@erasmusmc.nl)

different screening scenarios and the relative impact of screening and treatment on breast cancer mortality. Moreover, the simulation results provided quantitative information about the harms and benefits of various screening strategies not examined in randomized clinical trials, and have been used by policy makers to inform decisions about breast cancer screening guidelines.<sup>3,5</sup>

The consistency of previous collaborative modeling research provides a level of evidence for cross-validation. However, none of the prior collaborative CISNET research by the Breast Working Group has included external model validation. The International Society for Pharmacoeconomics and Outcomes Research in collaboration with the Society for Medical Decision Making (ISPOR-SMDM) recommends external model validation as part of good modeling practices, where external model validation is defined as, “the comparison of model predictions to observed event data not used in model development”.<sup>1</sup> The purpose of this paper is to conduct an external validation and compare CISNET breast cancer incidence and mortality predictions to observed clinical trial results of mammography screening from ages 40 to 49 years.

To date, the model parameters have been primarily developed based on US data on breast cancer epidemiology, screening, treatment, and population demographics.<sup>6</sup> Outcomes of our simulations indicated that offering screening to women in their fifties results in a more favorable ratio of benefits and harms than offering screening to women in their forties.<sup>3,7</sup> This difference in the benefits and harms between these age groups, corresponds to the available evidence of screening women aged 50 years and older<sup>8</sup> and the uncertainty about

screening women in their forties, considering the inconclusive evidence from fewer studies, and the different guidelines for this younger age group.<sup>5,9,10</sup> Given the high prevalence of dense breast tissue, faster growing tumors, and the inferior sensitivity of mammography in these younger women,<sup>11–13</sup> it is important to validate the models for the effectiveness of screening in the forties. The UK ‘Age’ trial is a well-documented<sup>14–20</sup> trial, investigating the effect of annually screening women from ages 40 to 49 years compared with no screening, and provided a unique opportunity to externally validate the CISNET breast cancer models for screening in the forties.

In this study, we present the first external validation performed by the CISNET breast cancer models that use different structures and assumptions about breast cancer natural history to project the impact of screening and treatment. We compare breast cancer incidence and mortality predictions to the observed results from the UK Age trial. The findings from this study are intended to inform CISNET model users as they can account for this information when considering and interpreting future model outcomes.

## Methods

The UK Age trial was the only randomized controlled trial designed specifically to investigate the effect of annual mammography screening from ages 40 to 49 years. Between October 1990 and September 1997, 160,836 women aged 40 to 41 years were randomly assigned in a ratio of 1:2 to either the intervention group or the control group. The 53,883 women in the intervention arm were offered annual screening by mammography, and the 106,953 women in the control arm received usual care (no screening). We collaborated with the Age trial investigators to obtain the observed de-identified data from the trial.

## Simulation Models

Five CISNET breast cancer models were included in this analysis: Model D (Dana-Farber), Model E (Erasmus), Model M (MD Anderson), Model S (Stanford), and Model W (Wisconsin-Harvard). These models have been developed independently within CISNET over the past 15 years and are described in detail elsewhere.<sup>21–25</sup> Briefly, women are born in a breast cancer-free stage, some women develop a tumor that may progress to a pre-clinical stage where it could be screen-detected in its pre-clinical sojourn time, or be diagnosed with breast

---

Department of Public Health, Erasmus Medical Center, Rotterdam, the Netherlands (JJV, NTV, HJD); Department of Oncology, Georgetown-Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington DC, USA (JSM); Department of Radiology, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA (ESB); Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA (MAE, OA, AT); Department of Radiology, School of Medicine, Stanford University, Stanford, CA, USA (CX, SKP); Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard Medical School Boston, Boston, MA, USA (HH, SJL); Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA (NKS); Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA (YL, JS); Department of cancer prevention, Wolfson Institute, Queen Mary University of London, London, UK (SMM). This work was supported by the National Institutes of Health under National Cancer Institute Grants U01CA152958 and U01CA199218.

**Table 1** Key Differences and Similarities between the CISNET Breast Models

| Model   | D   | E   | M   | S  | W   |
|---|---|---|---|--|---|
| <b>Model type</b>                             | Analytic, Parallel universe                       | Simulation, Parallel universe                                     | Bayesian, Parallel universe                       | Simulation, Parallel universe                        | Simulation, Parallel universe                     |
| <b>Natural history modeled as</b>             | State-transition                                  | Continuous tumor growth   | Bayesian model                                    | Continuous tumor growth                              | Continuous tumor growth                           |
| <b>Tumor inception</b>                        | Start of the sojourn time                         | Prior to start of sojourn time                                    | N/A   | Prior to start of sojourn time                       | Start of the sojourn time                         |
| <b>DCIS included</b>                          | Since 2014  | Yes   | Yes   | No   | Yes   |
| <b>Tumor ER status</b>                        | Yes   | Yes   | Yes   | Yes  | Yes   |
| <b>Screen detection depends on</b>            | Modality, age, density, frequency                 | Tumor size, modality, age, density, frequency                     | Modality, age, frequency                          | Tumor size, ER status, age, hormone repl., frequency | Tumor size, modality, age, density, frequency     |
| <b>Screening benefit</b>                      | Stage shift                                       | Detection at smaller tumor size                                   | Stage shift, beyond stage shift                   | Stage shift, smaller tumor size                      | Younger age, smaller tumor size                   |
| <b>Estimation of over diagnosis</b>           | Difference screen and no-screen                   | Difference screen and no-screen                                   | Difference screen and no-screen                   | Difference screen and no-screen                      | Difference screen and no-screen                   |
| <b>Treatment benefit</b>                      | Hazard reduction                                  | Cure fraction, larger fatal diameter                              | Cure fraction, hazard reduction                   | Hazard reduction, non-proportional                   | Cure fraction                                     |
| <b>Death from breast cancer determined by</b> | Survival from BC < survival other cause mortality | Fatal diameter, survival from BC < survival other cause mortality | Survival from BC < survival other cause mortality | Survival from BC < survival other cause mortality    | Survival from BC < survival other cause mortality |

**Model type.** Analytic: Analytical approach to estimate the impact of mammography screening and treatment on the incidence and mortality of breast cancer. Simulation: Stochastic simulation is based on the Monte Carlo method and the use of random numbers. Bayesian: The model does not include a natural history and estimates prior probability distributions for all unknown parameters. Parallel universe: Screening and treatment is modeled in a parallel universe, implying that the same population is simulated twice: once to determine the impact of breast cancer without screening, and once to determine the impact of breast cancer with screening. Breast cancer natural history and breast cancer death. ER: Onset and progression of breast cancer is different for estrogen receptor-positive and -negative tumors. Tumor stage transition: Tumor progression is modeled as transitions between different stages of breast cancer. Continuous tumor growth: Tumors grow continuously after tumor onset. Death from breast cancer: Once diagnosed with breast cancer, a survival until breast cancer death is competing with other-cause mortality survival. That is, breast cancer death occurs only if the patient does not die from other causes.

**Screening and Treatment.** Sensitivity: Sensitivity can be used directly or indirectly (e.g., when translated to tumor size). Over diagnosis: The detection and diagnosis of a condition that would not go on to cause symptoms or breast cancer death in a woman's lifetime. Hazard reduction: Reduction in breast cancer mortality hazard, calculated by 1 minus the hazard ratio for the different treatment regimes. Cure fraction: If hazard rate reduction is not a model input, it is translated into a cure fraction.

**Table 2** Model Inputs Used for the Age Trial Simulation

| Model Input                             | Description   | Source              |
|---|---|---------------------|
| <b>Population demographics</b>          |   |                     |
| Birth cohort                            | Birth years of women participating in the Age trial   | Age trial           |
| Life years                              | Number of life years by trial arm by age  | Age trial           |
| <b>Natural history of breast cancer</b> |   |                     |
| Incidence                               | Control arm incidence (incidence in the absence of screening)   | Age trial           |
| Tumor onset                             | The moment tumors start to grow (tumor inception)   | CISNET <sup>a</sup> |
| Sojourn time                            | Time between when a cancer is first screen-detectable and cancer diagnosis in the absence of screening                    | CISNET <sup>a</sup> |
| Tumor progression                       | Tumor growth, tumor progression and regression affect tumor sojourn times and breast cancer survival                      | CISNET <sup>a</sup> |
| Estrogen receptor distribution          | Age-specific ER positive and ER negative distributions  | UK <sup>b</sup>     |
| <b>Breast cancer screening</b>          |   |                     |
| Attendance                              | Adherence to annual screening in the intervention arm   | Age trial           |
| Sensitivity                             | Probability that the screen will be positive among women with breast cancer by age, screening round (first v. subsequent) | Age trial           |
| Mammography                             | Two-view mammography for first screens; one-view mammography for all subsequent screens                                   | Age trial           |
| <b>Breast cancer treatment</b>          |   |                     |
| Treatment dissemination                 | Breast cancer treatment by age, stage, and ER-status  | BASO <sup>c</sup>   |
| Effectiveness                           | Hazard reduction breast cancer mortality by age and ER-status   | EBCTCG <sup>d</sup> |
| <b>Breast cancer survival</b>           |   |                     |
| Survival                                | Breast cancer survival by age, stage, and ER-status   | CISNET <sup>e</sup> |
| Other-cause mortality                   | Probability of dying from causes other than breast cancer   | U.K. <sup>f</sup>   |

<sup>a</sup>Tumor onset, sojourn time, and tumor progression are model-specific parameters. These, and other model-specific assumptions about breast cancer natural history are described elsewhere.<sup>6, 21-25</sup>

<sup>b</sup>Estrogen receptor status comes from observed UK data.<sup>26</sup>

<sup>c</sup>The treatment dissemination was derived from BASO reports<sup>26</sup> published by the NHSBSP.

<sup>d</sup>Treatment effectiveness/hazard reduction for breast cancer death was published by the Early Breast Cancer Trialists Collaborative Group (EBCTCG), which included the UK trials.<sup>27</sup>

<sup>e</sup>Breast cancer survival by age and ER status from the UK is not available for the time period of the trial; the existing survival in the models based on USA data was used.

<sup>f</sup>Other-cause mortality was taken from the Human Mortality Database<sup>30</sup> with breast cancer deaths removed.

cancer due to clinical symptoms. Once diagnosed with breast cancer, women receive age-, stage-, and biomarker-specific treatment. Breast cancer incidence and mortality projections depend on age, start and stopping ages of screening, screening frequency, mammography screening performance, stage at diagnosis, estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) status of the tumor, breast cancer treatment, and factors related to the natural history of breast cancer (Tables 1 and 2). However, since the Age trial did not collect HER2 status, the models did not simulate HER2-specific molecular subtypes of breast cancer. The models adopt a ‘parallel universe’ approach; the same population of women is simulated twice: in one scenario women were invited to annual screening in their forties (intervention group), and in the second scenario women did not receive any screening in their forties (control group).

**Table 3** Number of Women in the Control and Intervention Groups

|           | No. of Women in the Control Arm | No. of Women in the Intervention Arm |
|-----------|---------------------------------|--------------------------------------|
| Age trial | 106,953                         | 53,883                               |
| Model D   | N/A <sup>a</sup>                | N/A                                  |
| Model E   | 10,000,000                      | 10,000,000                           |
| Model M   | 4,000,000                       | 4,000,000                            |
| Model S   | 5,000,000                       | 5,000,000                            |
| Model W   | 2,000,000                       | 2,000,000                            |

All models simulated at least about 20 times as many women in the control group and 40 times as many women as in the intervention group. The number of women simulated was selected by each model to balance feasibility of simulation time with a model output that yields relatively smooth incidence and mortality curves.

<sup>a</sup>Model D uses entirely analytical formulations to evaluate the impact of screening and treatment on breast cancer incidence and mortality; i.e., the number of women simulated does not apply to Model D.

As summarized in Table 1, the models differ in the ways they approximate unobservable events in the natural history of breast cancer. In model D, tumors progress via discrete state transitions,<sup>23</sup> models E, S and W have continuous tumor growth,<sup>21,22,25</sup> and model M uses Bayesian simulation<sup>24</sup> and does not have a natural history component. In models D and W, tumors are technically screen-detectable from tumor inception. Models E and S start simulating tumors at small tumor sizes, prior to the start of the sojourn time, when tumors are not yet screen-detectable by film or digital mammography. Screening benefit in models D and M is modeled as a stage shift to earlier stage breast cancer, with the latter model including an additional benefit of screening beyond stage shift. The benefits of screening in models E, S, and W are simulated by the detection of tumors at smaller sizes than at clinical diagnosis in the absence of screening (Table 1).

### Model Inputs

The Age trial data that the CISNET models obtained included control arm incidence in the absence of screening, mammography screening performance, screening attendance patterns, and demographic data such as life years and the distribution of birth years of women participating in the trial (Table 2). In the Age trial, data were not collected for breast cancer treatment. To fill this gap, we modeled the breast cancer treatment dissemination between 1991 and 2006, the intervention period of the trial, based on reports from the British Association of Surgical Oncology.<sup>26</sup> The effectiveness of breast cancer treatment was taken from analyses by the Early Breast Cancer Trialists' Collaborative Group (EBCTCG), which included trials conducted in the UK.<sup>27</sup> Model parameters related to the natural history of breast cancer, such as tumor onset and tumor growth, were based on the original CISNET parameters and no calibration was performed to the results from the Age trial.

### Simulation of the Age Trial

The women who participated in the Age trial were born between 1950 and 1957; therefore, we simulated the 1950 to 1957 birth cohort. In the trial, two-thirds of women aged 40 to 41 years were randomized to the control group and were not invited to any screening in their forties. The models simulated 2 to 10 million women in each arm of the trial, as they were not limited by practical issues concerning invitations and the number of women who can be included in the simulation of the trial (Table 3). Any

unscheduled screening in the control group was primarily a consequence of clinical symptoms and not because of routine screening;<sup>17</sup> therefore, we did not model explicitly such screening contamination in the control group.

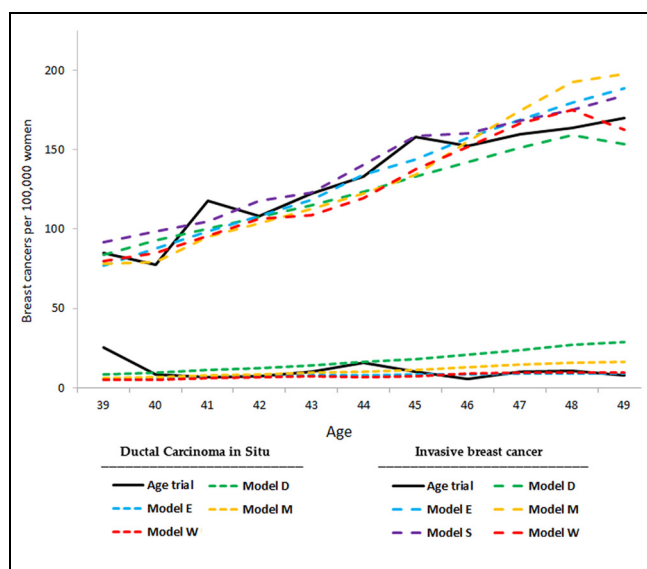
We used the control arm incidence as model input for a baseline projection of breast cancer incidence in the absence of screening. The models then overlaid the screening parameters according to the observed screening attendance patterns of the 53,883 women in the intervention group of the Age trial.<sup>18</sup> The percentage uptake of invitations increased by screening round, whereas the absolute number of invitations sent to women in the trial decreased by almost 50% near the end of the intervention period and, consequently, the absolute number of women who were screened decreased as well.<sup>18</sup> The models accounted for this by simulating the decrease in the number of women who were screened by age. The first analog mammogram in the trial included 2 views, and all subsequent mammograms were single-view, similar to the standard practice in the UK at the time of the trial. Screen detection of pre-clinical breast cancer was modeled based on observed sensitivity data published by the trial investigators.<sup>16</sup>

The UK treatment dissemination developed for this project indicated whether a breast cancer is treated with hormone therapy and/or chemotherapy after surgical removal of the tumor. Overall, ER-positive breast cancers were primarily treated with hormone therapy, and ER-negative breast cancers with chemotherapy. Since, the trial did not collect HER2 status, and Trastuzumab (Herceptin) was not yet disseminated in the UK at the time of the trial, it was not included in the treatment regimens.

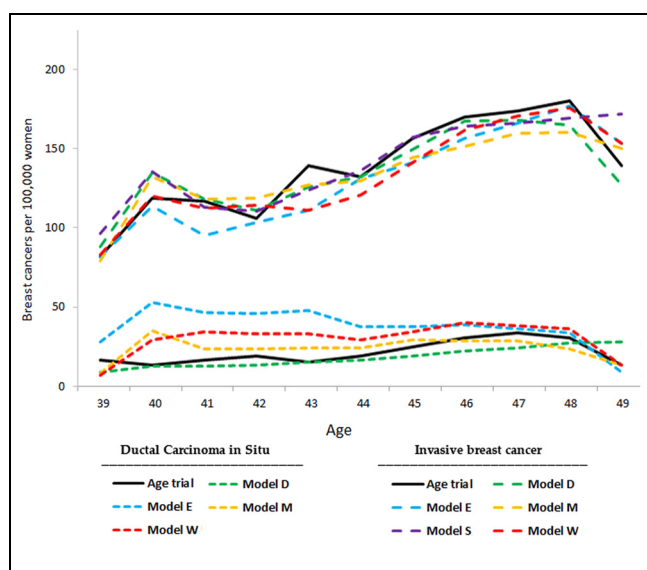
### Analysis

Model predictions were compared to breast cancer incidence and mortality observations from the Age trial by arm without calibrating the natural history parameters of the models to the trial. In addition, we compared the number of mammograms in the intervention group to that of the Age trial to investigate whether any differences in model predictions were related to variations in the number of mammograms.

We compared model outcomes with those from the trial at the 10-year and 17-year follow-up, corresponding to the most recent analysis by the Age trial investigators.<sup>15</sup> The trial used 'incidence-based mortality' to measure the effect of screening and treatment on breast cancer mortality. This implies only counting cancer deaths that originated from cancers diagnosed during the



**Figure 1** Control group breast cancer incidence (ductal carcinoma in situ [DCIS] and invasive separately) per 100,000 women, compared with the Age trial.



**Figure 2** Intervention group breast cancer incidence (ductal carcinoma in situ [DCIS] and invasive separate) per 100,000 women, compared with the Age trial.

intervention phase of the trial (ages 40 to 49 years). This is necessary because all women from both the intervention and control group ‘rolled’ into the national UK breast cancer screening program at age 50 and were invited to screening once every 3 years. For example, if at

age 54 years, there would be fewer breast cancer deaths among women randomized to the intervention group than among the women in the control group, one could conclude that the intervention of annual screening in the forties effectively reduced breast cancer mortality at age 54. However, because all women ‘rolled’ into the national screening program at age 50 years, it may be the case that the breast cancer deaths prevented at age 54 were actually from breast cancers diagnosed by screening at age 50 as part of the national program and not by the trial’s annual screening intervention in the forties. Therefore, the trial and the models only used breast cancer deaths from cancers diagnosed during the intervention phase to measure the effect of annual screening in the forties on breast cancer mortality.

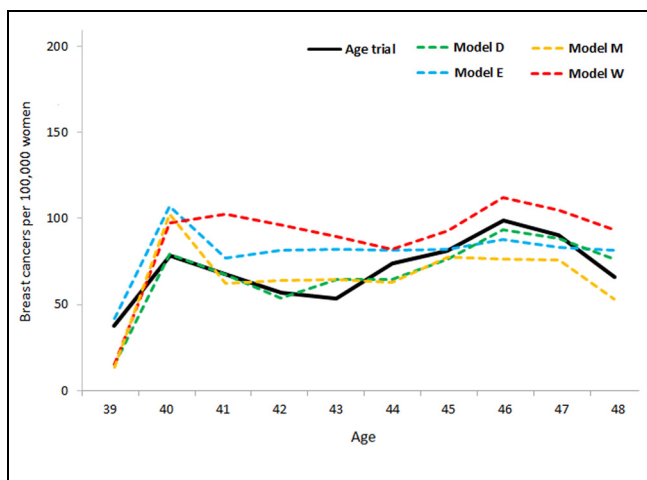
The confidence intervals associated with the mortality reduction observed in the Age trial at the 10-year and 17-year follow-up are useful as these are mainly influenced by the finite number of women included in the trial. The CISNET models have not included confidence intervals on their results given the millions of women simulated per trial arm. The model estimates will have a negligible range, given that the model outcomes are based on simulations of millions of women, each with varying combinations of variables constituting the life history, and sampled across the distribution of each variable. However, the model results do have uncertainty due to assumptions about unobservable parameters and structural uncertainties that are addressed. The use of multiple models provides a range of results that captures this structural uncertainty and could be considered to provide information comparable conceptually to a confidence interval.

## Results

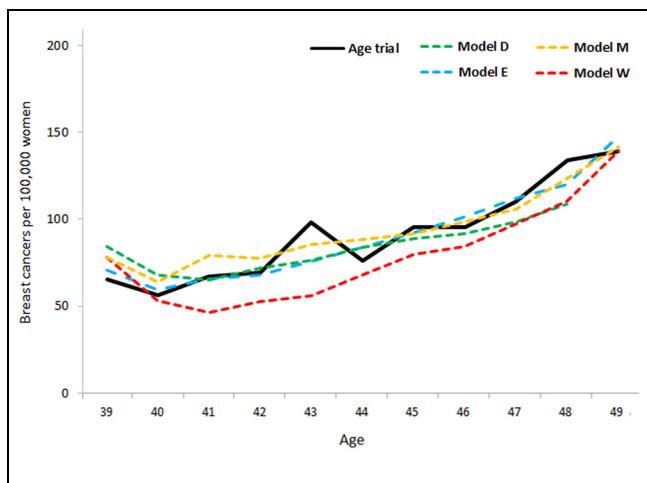
### Breast Cancer Incidence

The average simulated invasive breast cancer incidence among women aged 40 to 49 years in the control arm was 131 per 100,000 women (range across models, 124 to 138) as compared with 132 observed in the Age trial (Figure 1). The modeled ductal carcinoma in situ (DCIS) incidence was 11 per 100,000 women on average (range across models, 7 to 17), and equivalent to the 11 per 100,000 observed in the Age trial.

The average number of mammograms per woman in the intervention arm of the simulated trial was 5.2 (range across models, 4.9 to 5.4) compared with 4.84 in the Age trial. Modeled invasive breast cancer incidence in the intervention arm increased by age and was an average of 135 per 100,000 among women aged 40 to 49 years



**Figure 3** Intervention group (screen-detected) breast cancer incidence per 100,000 women. Screening ceased at age 48 years in the Age trial.



**Figure 4** Intervention group (clinically diagnosed) breast cancer incidence per 100,000 women.

(range across models, 131 to 141). This is consistent with the pattern for the 139 invasive breast cancers diagnosed per 100,000 women in the trial (Figure 2). DCIS intervention arm incidence varied more across the models (range, 18 to 38) and with 27 diagnoses on average, higher than the 21 DCIS diagnoses per 100,000 women in the trial. Models with continuous tumor growth (Models E and W) and models with tumor inception before the start of the tumor's sojourn time (Model E) tended to have the highest incidence of screen-detected DCIS.

**Table 4** Sensitivity of Screening in the Age Trial and in the Models

|                      | First Screen<br>(Two-view<br>Mammography) | Subsequent Screens<br>(Single-view<br>Mammography) |
|----------------------|---|--|
| Age trial            | 73.6                                      | 55.2   |
| Model D              | 73.6                                      | 55.3   |
| Model E              | 72.5                                      | 55.7   |
| Model M <sup>a</sup> | -   | -  |
| Model S              | 75.5                                      | 59.0   |
| Model W              | 67.7                                      | 59.6   |

<sup>a</sup>Model M is a Bayesian without a natural history part and a woman's disease status is unknown. As a result, sensitivity is not applicable.

Model M simulates screen- and clinically detected incidences without knowing the true disease status.

Sensitivity of screening and screen detection is modeled differently in various models. In the continuous tumor growth models E, S, and W screen detection of tumors is simulated by transforming sensitivity to a threshold tumor size at which tumors can be screen detected. On the other hand, model D uses sensitivity of screening by simulating a shift to a less-advanced stage of breast cancer.

Both the model results and the observed Age trial data included a small peak (Figure 3) at age 40 years in screen-detected breast cancers due to the detection of (prevalent) cases on the first mammogram, the only 2-view mammogram in the trial with better sensitivity than subsequent screens (Table 4). This was the only age during the trial at which the rate of screen-detected cancers was higher than the rate of clinically diagnosed cancers in the intervention group (Figure 4). The average rate of screen-detected DCIS and invasive breast cancers in the intervention arm in the age range 40 to 49 years was 69 per 100,000 women in the Age trial as compared with the models' average of 75 (range, 63 to 89). The rate of clinically diagnosed cases (DCIS and invasive breast cancers) in the intervention arm was 97 in the trial and 93 in the models (range, 82 to 99). Regardless of mode of detection, the rate of breast cancers diagnosed in the intervention arm between ages 40 - 49 was 161 per 100,000 women on average (range across models, 154 to 169) and similar to 162 in the Age trial.

### Breast Cancer Mortality

Among breast cancers diagnosed between ages 40 to 49 years, the Age trial found a total of 83 breast cancer deaths in the first 10 years of follow-up in the intervention arm (16 breast cancer deaths per 100,000 women) and 219 breast cancer deaths in the control arm (21 per 100,000 women). At the 10-year follow-up, the rate of

**Table 5** Breast Cancer Mortality Outcomes at the 10-Year Follow-Up

|                      | Mammograms per Woman | Breast Cancer Deaths per 100,000 Women |               | Rate Ratio Breast Cancer Deaths | Breast Cancer <sup>a</sup> Mortality Reduction |
|----------------------|----------------------|--|---------------|---------------------------------|--|
|                      |                      | Intervention Group                     | Control Group |                                 |  |
| <b>Age Trial</b>     | 4.84                 | 16                                     | 21            | 0.75                            | 25% (95% CI, 3% to 42%)                        |
| <b>Model Average</b> | 5.23                 | 19                                     | 23            | 0.85                            | 15.3% (range, 13% to 17%)                      |
| Model D              | 5.30                 | 17                                     | 20            | 0.83                            | 17.0%  |
| Model E              | 4.90                 | 20                                     | 25            | 0.83                            | 16.9%  |
| Model M              | 5.43                 | 20                                     | 23            | 0.86                            | 13.6%  |
| Model S              | 5.29                 | 22                                     | 25            | 0.87                            | 13.2%  |
| Model W              | 5.23                 | 19                                     | 22            | 0.84                            | 16.0%  |

<sup>a</sup>Age trial measured the effect of annual screening of women aged 40 to 49 years on breast cancer mortality. Therefore, the trial and the simulation models excluded breast cancer deaths that occurred in women diagnosed with breast cancer before age 40 and after age 49 years.

**Table 6** Breast Cancer Mortality Outcomes at the 17-Year Follow-Up<sup>a</sup>

|                      | Mammograms per Woman | Breast Cancer Deaths per 100,000 Women |               | Rate Ratio Breast Cancer Deaths | Breast Cancer <sup>a</sup> Mortality Reduction |
|----------------------|----------------------|--|---------------|---------------------------------|--|
|                      |                      | Intervention Group                     | Control Group |                                 |  |
| <b>Age Trial</b>     | 4.84                 | 19                                     | 22            | 0.88                            | 12% (95% CI, -4 to 26%)                        |
| <b>Model Average</b> | 5.23                 | 20                                     | 23            | 0.87                            | 13.2% (range, 10% to 17%)                      |
| Model D              | 5.30                 | 20                                     | 22            | 0.90                            | 9.7%   |
| Model E              | 4.90                 | 18                                     | 22            | 0.83                            | 17.1%  |
| Model M              | 5.43                 | 20                                     | 24            | 0.85                            | 15.2%  |
| Model S              | 5.29                 | 21                                     | 24            | 0.89                            | 11.0%  |
| Model W              | 5.23                 | 18                                     | 21            | 0.86                            | 13.7%  |

<sup>a</sup>The Age trial measured the effect of annual screening of women aged 40 to 49 on breast cancer mortality. Therefore, the trial and the simulation models excluded breast cancer deaths that occurred in women diagnosed with breast cancer before age 40 and after age 49.

breast cancer deaths per 100,000 women predicted by the models was 20 on average (range across models, 17 to 22) in the intervention arm, and 23 (range across models, 20 to 25) in the control arm (Table 5). The number of breast cancer deaths predicted by the different models was consistently somewhat higher in both arms than in the trial.

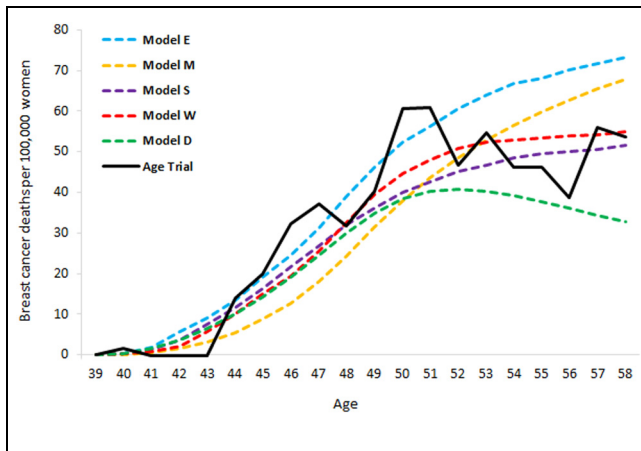
On average, the modeled breast cancer mortality reduction due to screening was 15% (range across models, 13% to 17%) at the 10-year follow-up v. 25% (95% CI, 3% to 42%) observed in the Age trial. At the 17-year follow-up, the models predicted 13% (range across models, 10% to 17%) breast cancer mortality reduction when restricted to breast cancer deaths that originated from breast cancers diagnosed during the intervention phase (incidence-based mortality) v. 12% (95% CI, -4% to 26%) observed in the trial (Table 6). The models with either tumor onset at small tumor sizes before the start of the sojourn time and on average slow tumor progression (Model E), or with tumor cure fractions for

treatment benefit (Models E, M and W) maintained their 10-year follow-up breast cancer mortality reduction predictions at the 17-year follow-up, whereas mortality reduction in the trial decreased. Similar to the Age trial, the models showed a turning point around age 50 years, where the increase in the cumulative number of breast cancer deaths averted started to diminish (Figure 5).

## Discussion

This is the first collaborative CISNET breast cancer study comparing model predictions to observed clinical trial results not used in the development of any model parameters. The results indicate that all 5 models estimate the long-term effect of annual screening between the ages of 40 to 49 years well within the observed confidence intervals of the UK Age trial. The impact of screening on breast cancer mortality was also internally consistent with individual model structures regarding the natural history of breast cancer.





**Figure 5** Cumulative breast cancer deaths averted per 100,000 women. Cumulative breast cancer deaths averted only using breast cancer deaths from cancers diagnosed in the intervention period per 100,000 women. Calculated by the rate of breast cancer deaths in the control group minus the rate of breast cancer deaths in the intervention group.

The ISPOR-SMDM Modeling Good Research Practices TaskForce-7<sup>1</sup> states that predictive and external validation are the strongest forms of model validation, as decision-makers can account for this information when considering model outcomes. In the past, the breast CISNET models have illustrated accurate predictions of molecular-subtype-specific and overall US breast cancer incidence and mortality trends.<sup>3,4,28</sup> This study extends these earlier cross-validations by independently estimating the observed results from a UK randomized controlled trial.

All models reproduced the trend in control group breast cancer incidence from ages 40 to 49 years, implying that the extant model structures and assumptions about the natural history of breast cancer in the absence of screening are reliable. Despite the intensive (annual) screening intervention, the models predicted more clinically diagnosed than screen-detected breast cancers in the intervention group. This was likely to be explained by the relatively low sensitivity of all subsequent single-view mammograms that followed after the more sensitive prevalent 2-view mammograms, and the decrease in the number of women screened by screening round in the trial.<sup>18</sup> Although the models utilized different mechanisms, such as a threshold tumor size (Models E, S, and W) or stage shift (Models D and M) to simulate screen detection of pre-clinical breast cancer, they were all able to accurately estimate the impact of screening from ages 40 to 49 years on invasive breast cancer incidence.

The effect of screening and treatment on breast cancer mortality was underestimated by all models at the 10-year follow-up as compared with the reduction observed in the Age trial. Because all models accurately predicted breast cancer incidence, and given that the underestimation of the mortality reduction was present across all models, it might be explained by a common model input not related to screening. Specifically, the derived UK treatment dissemination may not represent the actual treatment received by women diagnosed with breast cancer in the trial. This is in line with the higher rate of breast cancer deaths predicted by the models in the control arm in the absence of screening.

After 10 years of follow-up, breast cancer mortality reduction observed in the trial decreased and lost significance, whereas most models predicted a fairly constant mortality reduction between the 10- and 17-year follow-ups. Previous analysis of the CISNET models<sup>29</sup> illustrated that Model D, with tumor inception at the start of the sojourn time, has fast tumor progression on average, and Model E, with tumor inception before the start of the sojourn time, has the slowest tumor progression on average. These individual model structures affect the pattern in breast cancer deaths averted after age 49 when screening ceased, because cancers diagnosed in the control arm caused breast cancer death at a younger age in Model D and at a later age in Model E. Consequently, mortality reduction due to screening was greater at later ages (between the 10- and 17-years follow-up) in Model E than in Model D. While the model structure of Model S is similar to that of Model E, Model S does not include DCIS, which implies no possible benefit in terms of mortality reduction from screen-detected DCIS. However, these otherwise screen-detected DCIS cases will likely be diagnosed as local stage small invasive tumors (size < 1 cm) in Model S with relatively high and similar survival as DCIS cases. Model W is unique in that it simulates tumors with a limited malignant potential.<sup>25</sup> This may have resulted in a substantial number of screen-detected tumors that did not cause breast cancer death during the 17-year follow-up. Consequently, Model W's mortality reduction decreased slightly after age 49 years despite their high rate of screen-detected cancers in the forties.

In summary, at the 10- and 17-year follow-up, the models reproduced the effects of annual screening in the forties on breast cancer mortality well within the trial's CIs.<sup>15</sup> In terms of model validation, it can be questioned what these model outcomes imply, as it is quite common to have relatively wide CIs in randomized trials on cancer screening. The wide CIs in the trial are partly due to the limited number of women included and breast cancer

deaths observed in the trial. The models' outcomes may be less sensitive to the number of women who are simulated because they simulated at least 2 million women in each arm of the trial, notwithstanding that the models are ultimately based on observed data as well.

The CISNET breast models used Age trial-specific model inputs and data sources applicable to the UK, but we can still draw a comparison between the outcomes of this study and published results from a recent collaborative modeling study on screening in the United States<sup>3</sup>. In the US study, we simulated annual screening from age 40 to 74 years and compared it to annual screening from age 50 to 74 years. This implies that the difference in breast cancer deaths averted between these 2 scenarios over the women's lifetime is because of annual screening in the forties. Like the results of this analysis, the outcomes indicated that Model M and E avert the most breast cancer deaths from annual screening in the forties followed by Models W, S, and D. In other words, the ranking of the models is consistent when applied in another country with different model inputs.

This study presented the first external comparison performed by multiple breast cancer simulation models applied in a different country and setting. A strength of this analysis is that we used detailed observed de-identified trial data as model inputs. Another important strength is that we performed an independent external validation<sup>1</sup> in which no model calibration was performed to ensure the credibility of the model outcomes.

Although the CISNET breast models used Age trial-specific model inputs and data sources applicable to the UK, there were several limitations in this analysis. The trial did not collect data on breast cancer molecular subtype and treatment; these were estimated based on UK data. It is possible that these data underestimated the actual treatment patterns of trial participants. That this is the case is suggested by the fact that all models had estimates for mortality reduction that were consistently lower than the point estimate from the trial. Moreover, when the models simulated the Age trial assuming all women received the most effective therapy available, the average model estimate was very close to trial result.<sup>3</sup> The lack of precision in modeling the treatment of women in the Age trial is likely to have contributed more to the differences between model and trial results than the screening and natural history components of the models. Another limitation was that the models did not explicitly simulate screening in the control arm because the reported amount of unscheduled screening was low, and primarily due to symptomatic reasons.<sup>17</sup> Although this may not affect our conclusions of the simulations, it is a limitation.

The quantitative information in this study demonstrated how well the models reproduced the effects of annual screening from ages 40 to 49 years on breast cancer incidence and mortality. In the future, the CISNET models could simulate the impact of what would have happened if 2-view digital mammography had been used for all screening examinations in the Age trial, or simulate the impact of different patterns of screening attendance, provide estimates on overdiagnosis, and estimate the lifetime effects of different screening programs offered to women in their forties. The demonstration that the models can reproduce observed external trial results should increase confidence in model results to inform policy decisions about breast cancer screening.

### Acknowledgment

The authors thank the Age trial investigators for providing de-identified Age trial data. The authors also acknowledge the contributions of Stephan W. Duffy to this project.

### References

1. Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Med Decis Making*. 2012;32:733–43.
2. Plevritis S, Munoz D, Kurian A, et al. Estimating breast cancer survival by molecular subtype in the absence of screening and adjuvant treatment. *Med Decis Making*. 2018;38(1S):32–43.
3. 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;164:215–25.
4. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med*. 2005 Oct 27;353:1784–92.
5. Siu AL, On behalf of the U. S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2016;164:279–96.
6. Mandelblatt JS, Near AM, Miglioretti DL, Munoz D, Sprague BL, Trentham-Dietz A, et al. Common model inputs used in CISNET collaborative breast cancer modeling. *Med Decis Making*. 2018;38(1S):9–23.
7. 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;156:609–17.
8. Independent U. K. Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet*. 2012;380:1778–86.
9. American College of O-G. Practice bulletin no. 122: Breast cancer screening. *Obstet Gynecol*. 2011;118:372–82.

10. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast cancer screening for women at average risk: 2015 Guideline update from the American Cancer Society. *JAMA*. 2015;314:1599–614.
11. Sprague BL, Gangnon RE, Burt V, et al. Prevalence of mammographically dense breasts in the United States. *J Natl Cancer Inst*. 2014;106.
12. Anders CK, Fan C, Parker JS, et al. Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes? *J Clin Oncol*. 2011;29:e18–20.
13. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med*. 2007;356:227–36.
14. 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;368:2053–60.
15. Moss SM, Wale C, Smith R, Evans A, Cuckle H and 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;16:1123–32.
16. Moss S, Thomas I, Evans A, Thomas B and Johns L, Trial Management G. Randomised controlled trial of mammographic screening in women from age 40: results of screening in the first 10 years. *Br J Cancer*. 2005;92:949–54.
17. Kingston N, Thomas I, Johns L and Moss S, Trial Management G. Assessing the amount of unscheduled screening ("contamination") in the control arm of the UK "Age" Trial. *Cancer Epidemiol Biomarkers Prev*. 2010;19:1132–6.
18. Johns LE and Moss SM, Trial Management G. Randomized controlled trial of mammographic screening from age 40 ('Age' trial): patterns of screening attendance. *J Med Screen*. 2010;17:37–43.
19. Johns LE and Moss SM, Age Trial Management G. False-positive results in the randomized controlled trial of mammographic screening from age 40 ("Age" trial). *Cancer Epidemiol Biomarkers Prev*. 2010;19:2758–64.
20. Moss S. A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. Trial Steering Group. *J Med Screen*. 1999;6:144–8.
21. van den Broek JJ, van Ravestein NT, Heijnsdijk EA and de Koning HJ. Simulating the impact of risk-based screening and treatment on breast cancer outcomes with MIS-CAN-Fadia. *Med Decis Making*. 2018;38(1S):32–43.
22. Munoz D, Xu C and Plevritis S. A molecular subtype-specific stochastic simulation model of US breast cancer incidence and mortality trends from 1975 to 2010. *Med Decis Making*. 2018;38(1S):89–98.
23. Lee SJ, Li X and Huang H. The Dana-Farber CISNET model for breast cancer screening strategies: An update. *Med Decis Making*. 2018;38(1S):44–53.
24. Huang X, Li Y, Song J and Berry DA. A Bayesian simulation model for breast cancer screening, incidence, treatment, and mortality. *Med Decis Making*. 2018;38(1S):78–88.
25. Alagoz O, Ergun MA, Cevik M, et al. The University of Wisconsin breast cancer epidemiology simulation model: An update. *Med Decis Making*. 2018;38(1S):99–111.
26. National Health Service Breast Screening Programme & British Association of Surgical Oncology. *Audit of screen detected breast cancers*. 2000–2007. Available at: <http://associationofbreastsurgery.org.uk/audit/nhs-breast-screening-programme-audit/>
27. Early Breast Cancer Trialists' Collaborative G, Peto R, Davies C, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet*. 2012;379:432–44.
28. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Int Med*. 2009;151:738–47.
29. van den Broek JJ, van Ravestein NT, Cevik M, et al. Comparing CISNET breast cancer models using the maximum clinical incidence reduction methodology. *Med Decis Making*. 2018;38(1S):112–125.