Supplementary material

# Model description

The core model is based on a previously published and validated Markov model, as described in Schousboe, Kerlikowske (1). The authors provided very detailed supplementary material describing the Markov model, incidence and mortality data, as well as relative risks, screening effects, costs, and utility assumptions. The core model specifies state transitions for healthy individuals toward non-invasive in situ cancers (DCIS) or invasive cancers, in one of three states: local, regional, or distant. DCIS is considered not to progress into invasive cancers, but acts as a risk factor for developing a second and invasive cancer. Cancer progression from local to regional or from regional to distant cancers is not modeled as such. These progressions are, however, reflected in the survival curves for each cancer state and are also reflected in the cost and treatment effects, as they are all calculated for the summary stage of the Surveillance, Epidemiology, and End Results Program (SEER) at time of diagnosis.

On top of the survival and mortality module, screening strategies are incorporated by allowing a stage shift from later to earlier disease states. The distributions, which describe whether an invasive cancer is localized, regional, or distant, shift when screening strategies are included. Screening thus does not prevent cancer, but allows an earlier diagnosis and longer survival in an earlier cancer stage. This model uses four distributions, which reflect cancer stages at the time of diagnosis for women who were not screened, for women with triennial, biennial, or annual screening visits. Schousboe, Kerlikowske (1) and Kerlikowske, Zhu (2) describe stage distributions for these screening strategies dependent on breast density. The original model by Schousboe, Kerlikowske (1) showed some irregularities in annual strategies compared to biennial strategies for the breast density-specific distributions, which would lead to a superiority of biennial strategies in all cases. This was considered implausible, as this superiority could not be found in the aggregated distributions in Schousboe, Kerlikowske (1) and the breast density-specific distributions in Kerlikowske, Zhu (2). For the annual strategy, we thus used the distributions from Kerlikowske, Zhu (2).

# Screening adherence and risk

Since the introduction of organized mammography programs, participation in screening has always been a sensitive subject for the success of these programs. The continuation of screening is often described as screening maintenance (3) or screening adherence (4). Both describe the participation in repeated mammography screenings according to an agreed schedule. In this simulation, we use the term “adherence” and assume that each adherence decision is randomly decided with each screening invitation. Accordingly, we do not model patterns of repeated non-adherence or correlated non-adherence deriving from “bad” experiences.

We added screening adherence and non-adherence as an additional module, which can be used to replicate 100% adherence for validation of other simulation models or allows the screening of non-adherence at an individual level. The module allows this by random decision-making of individuals on their random path through the screening module. At each screening event, a random number is drawn and compared with the individual’s average probability of adhering or not adhering. By drawing random numbers at each event, a random adherence path along the screening path is produced, which reproduces the individual’s average adherence to screening.

### Positive risk-dependent adherence

There is evidence that these forms of adherence are linked to the perceived risk of every woman. Although there is no agreement yet about the nature of the link, three hypotheses are being discussed and supported by empirical evidence. The first hypothesis is that women with higher perceived risk are more likely to adhere to screening, and the perception of risk depends on observable risk factors such as familial risk, earlier atypical biopsy findings, or high breast density.

McCaul, Branstetter (5) conducted a meta-analysis of 19 studies. They found a significant positive, but small, effect of higher levels of perceived risk on mammography adherence. Three years later, McCaul and Tulloch (6) did another systematic review on factors influencing cancer screening decisions. They confirmed the earlier finding from the meta-analysis and stated that risk perception is influenced by family risk. Family risk may, however, influence screening adherence over other paths, for example by prompting a physician recommendation. A later meta-analysis by Katapodi, Lee (7) confirmed the influence of perceived risk on screening adherence and found that women often underestimate their perception. This optimism was not found in women with a family history and thus a higher perception of their own hereditary risk. Hay, McCaul (8) showed in a meta-analysis that breast cancer worry influences screening behavior, and higher levels of worry are very uncommon, but increase the likelihood of screening. The latest review by Walker, Chiarelli (9) aimed at supporting the earlier studies and found that there is a weak positive association in women with familial risk, the perception of higher risk, and screening adherence. From these studies, we derive the first scenario: women with higher risk have higher perceived risk and are more likely to adhere to screening recommendations.

### Negative risk-dependent adherence

The second cluster of studies found the opposite association between risk and adherence: high perceived risk may lead to psychological distress, and any form of psychological distress causes non-adherence to mammography screening.

Kash, Holland (10) assessed levels of psychological distress and anxiety in women at high risk of developing breast cancer because of family history. They found that women experiencing barriers to accessing screening had higher levels of psychological distress and a greater need for counseling. At the same time, high levels of anxiety were connected with lower attendance rates at clinical breast examination and also lower rates of self-examination. Schwartz, Taylor (11) and Schwartz, Taylor (12) focused on women under psychological distress and found negative associations with screening adherence in retrospective (11) and prospective data (12). However, Schwartz, Taylor (12) described how they did not find an association when assessing cancer-specific distress, but only with moderate “worry”. They explain that much of the controversy derives from the different definitions and assessments of psychological distress and worry. Castello, Prieto (13) assessed the association between mammographic breast density and adherence to screening guidelines in a cross-sectional study of 3,584 women in Spain. They found that compliance with guidelines was higher in women with lower breast density, and thus recommend specific guidance for women with higher breast density. However, they did not control for whether higher breast density leads to more psychological distress or offer an alternative explanation for how breast density affects screening adherence.

There are no review studies or meta-analyses supporting these findings as in the first hypothesis, but the discussion is on-going, and recent empirical studies still occasionally support the negative effect of psychological distress or worry on adherence. That is why this paper does not solely rely on the first scenario, but also allows a second scenario with the opposite effect. The second scenario can thus be described as: women with higher risk are more likely to experience psychological distress and are less likely to adhere to screening.

### Curvilinear risk-dependent adherence

A third group of studies exists aiming to bridge the divide between the first and the second hypotheses. These studies found that moderate levels of psychological distress or worry lead to increased compliance, but low or high levels of psychological distress have detrimental effects. Psychological distress thus influences adherence in an inverted U-shaped curve, which is often called a curvilinear relationship.

Lerman, Caputo (14) conducted telephone interviews to identify screening behavior and barriers to participation. They found that having a higher perceived risk of breast cancer, for example when having a family history, has a positive association, but anxiety about screening reduced the likelihood of repeated participation. The perception of having a higher risk thus increases the motivation to adhere to screening but, if the higher risk or the mammography screening leads to anxiety, the likelihood of adhering to repeated screenings decreases. Hailey (15) reviewed the conflicting findings of positive and negative effects of perceived risk and concluded that anxiety influences screening behavior in an inverted U-shaped curve. Moderate levels of anxiety accordingly would increase adherence, but too low or too high levels have detrimental effects on adherence. Andersen, Smith (16) aimed to find empiric evidence for the curvilinear relationship and interviewed women to prove or disprove the hypothesis. They found evidence for the inverted U-shaped curve and also that the U-shape differs in average and high risk women. They concluded that a “barrier effect” of severe worry exists, and this effect is more pronounced in women at high risk. Lemon, Zapka (17) found a similar pattern in a sample of women aged over 50 years and concluded that it might be proof of a curvilinear relationship between breast cancer worry and mammography.

These studies thus suggest that elevated perceived risk of developing breast cancer leads to complex psychological responses and may have positive effects on screening behavior, but can also act as a barrier to participation. The third scenario is thus: increased perceived risk or worry of developing breast cancer affects screening adherence in an inverted U-shaped curve. Moderate risk levels thus increase adherence, whereas low or high risk levels decrease the likelihood of adherence.

### Technical implementation

In this simulation study, we introduce an adherence variable into a validated simulation model and create three scenarios of positive, negative, or curvilinear relationships between risk levels and adherence behavior to analyze how these scenarios affect the cost-effectiveness of risk-stratified screening strategies. The risk levels are reflected by four risk factors, age, breast density, family history in a first-degree relative, and having had a previous breast biopsy. These three risk factors can be used to calculate a risk score based on the Breast Cancer Surveillance Consortium (BCSC) risk calculator tool developed by Tice, Cummings (18). We assume that, at the moment of the first screening, each woman is confronted with her risk level and is assigned a probability of adhering to screening based on risk score.

For the age group of 50-year-old women, there are 16 combinations of risk factors, which translates to a spread between 0.41% to 4.65% risk of developing breast cancer within 10 years. We fitted logarithmic functions to represent the positive and negative relationship and quadratic functions to represent curvilinear relationships. The logarithmic function for positive or negative associations was chosen to best represent the risk distribution in the population. The quadratic function for the curvilinear function was fitted to best represent the curvilinear nature reported by Andersen, Smith (16). All functions are fitted to represent an effect size of 19% between risk perception and adherence, as reported by Katapodi, Lee (7), and an average adherence rate of 72.4%, as reported by Centers for Disease Control and Prevention (19). The functional forms are:

Table S1 presents the three adherence scenarios, the corresponding risk levels, and the expected participation rate assuming the prevalence of risk factors as in Schousboe, Kerlikowske (1). Breast density levels use the categorization of the Breast Imaging and Data System (BI-RADS)

Table S1: Adherence level scenarios and risk score

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Breast density (in BI-RADS levels) | Family history (y/n) | Previous biopsy (y/n) | 10-year risk of developing invasive breast cancer (in %) (18) | Adherence probability (%) |
| Positive association | Negative association | Curvilinear association |
| 1 | 0 | 0 | 0.41 | 62.91 | 81.87 | 65.00 |
| 1 | 0 | 1 | 0.62 | 66.14 | 78.65 | 67.44 |
| 1 | 1 | 0 | 0.78 | 67.94 | 76.86 | 69.10 |
| 2 | 0 | 0 | 0.85 | 68.61 | 76.19 | 69.77 |
| 1 | 1 | 1 | 1.17 | 71.11 | 73.70 | 72.43 |
| 2 | 0 | 1 | 1.28 | 71.81 | 72.99 | 73.18 |
| 3 | 0 | 0 | 1.37 | 72.34 | 72.46 | 73.74 |
| 2 | 1 | 0 | 1.60 | 73.56 | 71.25 | 74.92 |
| 4 | 0 | 0 | 1.67 | 73.89 | 70.92 | 75.21 |
| 3 | 0 | 1 | 2.05 | 75.49 | 69.32 | 76.22 |
| 2 | 1 | 1 | 2.39 | 76.69 | 68.12 | 76.30 |
| 4 | 0 | 1 | 2.50 | 77.05 | 67.77 | 76.16 |
| 3 | 1 | 0 | 2.57 | 77.26 | 67.56 | 76.03 |
| 4 | 1 | 0 | 3.13 | 78.80 | 66.02 | 73.82 |
| 3 | 1 | 1 | 3.83 | 80.38 | 64.45 | 68.12 |
| 4 | 1 | 1 | 4.65 | 81.90 | 62.93 | 57.29 |
| Expected participation rate | 72.50 | 72.30 | 72.99 |

In order to test these estimates in the univariate and probabilistic sensitivity analysis, the values were randomly changed to reflect ranges from different studies, as reported by Jacklyn, Glasziou (20). They reported adherence to breast cancer screening as reported in eight different trials with ranges from 65% to 100%. In order to reflect these in the average participation, the adherence probabilities are varied between 90% and 135% around the mean value.

## Relative risk and prevalence of breast density levels

Table S2: Relative risk due to breast density

|  |  |
| --- | --- |
| Age (years) | Relative risk of breast density in BI-RADS categories |
| BI-RADS 1 | BI-RADS 2 | BI-RADS 3 | BI-RADS 4 |
| 40–49 | 0.351 | 0.730 | 1.131 | 1.468 |
| 50–59 | 0.388 | 0.807 | 1.251 | 1.623 |
| 60–64 | 0.400 | 0.832 | 1.291 | 1.675 |
| 65–69 | 0.581 | 0.885 | 1.228 | 1.283 |
| 70+ | 0.600 | 0.914 | 1.268 | 1.325 |

Table S3: Age-specific BI-RADS distribution and probability of reducing density in 10 years

|  |  |
| --- | --- |
| Age (years) | Distribution of BI-RADS (probability of reducing density) |
| BI-RADS 1 | BI-RADS 2 | BI-RADS 3 | BI-RADS 4 |
| 40–49 | 4.4% | 35.3% (9.6%) | 46.8% (31.4%) | 13.5% (52.6%) |
| 50–59 | 7.8% | 46.6% (3.4%) | 39.2% (11.5%) | 6.4% (18.8%) |
| 60–69 | 9.4% | 49.5% (5.1%) | 35.9% (19.2%) | 5.2% (34.6%) |
| 70+ | 11.9% | 53.9% | 30.8% | 3.4% |

## Utility and cost input parameters

We used cost and utility parameters as described by Schousboe, Kerlikowske (1). All monetary values are in 2016 USD.

Table S4: QALY utility parameters

|  |  |  |
| --- | --- | --- |
| Description | Value | Source |
| QALY weight |  |  |
| …until 49 years | 0.859 | (1, 21) |
| …until 59 years | 0.845 | (1, 21) |
| …until 69 years | 0.812 | (1, 21) |
| …until 79 years | 0.788 | (1, 21) |
| …over 80 years | 0.762 | (1, 21) |
| QALY loss |  |  |
| …in first year of in situ | 10% | (1, 21) |
| …in following years of in situ | 0% | (1, 21) |
| …in first year of local cancer | 15% | (1, 21) |
| …in following years of local cancer | 2% | (1, 21) |
| …in first year of regional cancer | 25% | (1, 21) |
| …in following years of regional cancer | 10% | (1, 21) |
| …in first year of distant cancer | 25% | (1, 21) |
| …in following years of distant cancer | 17% | (1, 21) |
| …from false-positive screening | 0.013 | (1) |

Table S5: Cost parameters in 2016 US$

|  |  |  |
| --- | --- | --- |
| Description | Value in 2016 US$ | Source |
| Mammography screening | 138 | (1) |
| Cost of false-positive results | 507 | (1, 22) |
| DCIS treatment |  |  |
| …initial care in first year | 11,371 | (1, 23) |
| …continuing care in following years | 994 | (1, 23, 24) |
| Localized cancer treatment |  |  |
| …initial care in first year | 14,970 | (1, 23) |
| …continuing care in following years | 708 | (1, 23, 24) |
| …terminal care in last year | 40,521 | (1, 23) |
| Regional cancer treatment |  |  |
| …initial care in first year | 28,304 | (1, 23) |
| …continuing care in following years | 4,103 | (1, 23, 24) |
| …terminal care in last year | 47,963 | (1, 23) |
| Distant cancer treatment |  |  |
| …initial care in first year | 43,714 | (1, 23) |
| …continuing care in following years | 12,044 | (1, 23, 24) |
| …terminal care in last year | 67,273 | (1, 23) |

# Validation

In order to provide a structured assessment of the model validation, the AdViSHE tool was used as a guide (25). The following section follows the four modules: conceptual model, input data, computerized model, operational validation.

## Conceptual model validation

The original model concept was derived and validated by Schousboe, Kerlikowske (1). The new concept added here, allowing adherence to influence screening behavior, was conceptualized by MA and checked to represent epidemiologic and medical context by AQ. In addition, the model concept has been presented in workshops and conferences and adopted to reflect expert opinion and feedback. As there are currently no empiric data to validate the true relation between risk and adherence, the uncertainty is represented using scenarios to describe positive, negative, curvilinear, or random relationships. The extent to which the relationship is varied in sensitivity analysis allows additional parameter uncertainty. With this approach, the uncertain nature is reflected until a data source for empiric validation is available in future research.

## Input data validation

Most of the input data could be accessed from the supplementary material of the original model (1). The validation of breast cancer incidence, screening sensitivity and specificity, relative risk, and mortality had already been conducted by Schousboe, Kerlikowske (1) and assessed to replicate real-world data from the SEER (26) and BSCS databases (27). The newly added adherence parameters were constructed by MA and checked by AQ. Adherence parameters were constructed to reflect real-world average adherence levels as derived from the SEER estimates (26) and varied in sensitivity analyses to reflect the bandwidth of adherence levels, as reported in meta-analyses of clinical trials (20).

## Computerized model validation

We used TreeAge Pro Healthcare for the implementation of the model (28). The computerized model was tested by testing the plausibility of the cancer incidence and mortality calculated with and without screening, the effects of the different screening intervals on incidence and mortality, the plausibility of integrating harmful effects or excluding them, allowing screening non-adherence or assuming full adherence under screening. These different modules were tested by checking Markov traces of individuals for average and extreme values. Individuals with very low or very high risk profiles were traced through the model using TreeAge Pro’s global matrix function, which produces event traces for each individual in every cycle. These traces can be exported using a spreadsheet software, such as MS Excel (29). These traces were calculated for individuals with extreme risk profiles under the assumption of no screening, routine screening, adherence-influenced screening, and risk-dependent adherence.

## Operational validation

We validated the operational function of the model in four steps. First, we predicted breast cancer incidence and mortality rates and compared them with the estimates from the original model and real-world data. In order to compare the estimates with the recent SEER estimates (26), we assumed that recent estimates were influenced by the current screening guidelines and screening participation. The current screening recommendation is biennial screening for women between 50 and 74 years, and the corresponding participation rate was 72.4% in 2010 (19).

Table S6: Cumulative lifetime (age 100 years) incidence and mortality by start age, under the assumption of 72.4% participation in current guidelines of biennial mammography between 50 and 74 years

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Incidence (%) |  |  | Mortality (%) |  |  |
| By start age (years) | Own calculation | SEER (26) | Schousboe, Kerlikowske (1) | Own calculation | SEER (26) | Schousboe, Kerlikowske (1) |
| 40 | 11.50 | 12.2 | 12.35 | 2.78 | 2.8 | 2.99 |
| 50 | 10.51 | 11.1 | 11.12 | 2.68 | 2.6 | 2.90 |
| 60 | 8.28 | 9.4 | 8.84 | 2.34 | 2.4 | 2.45 |
| 70 | 5.18 | 6.7 | 5.99 | 1.73 | 2.0 | 2.02 |

Second, we compared the relative risks of breast density in our model with other estimates. The relative risks was calculated with BI-RADS 2 as the base case. For the calculation, cumulative incidence numbers were compared for models with fixed breast density levels. These models calculated incidence numbers with a model start age of 40 years and lifetime (up to 100 years) time horizon.

Table S7: Model predicted relative risk of breast density

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Age 40–65 years |  |  | Age 65+ years |  |  |
| BI-RADS level | Own calculation | Schousboe, Kerlikowske (1) | Tice, Cummings (18) | Own calculation | Schousboe, Kerlikowske (1) | Tice, Cummings (18) |
| 1 | 0.48 | 0.48 | 0.49 | 0.69 | 0.67 | 0.66 |
| 2 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| 3 | 1.52 | 1.53 | 1.55 | 1.32 | 1.37 | 1.39 |
| 4 | 1.94 | 1.97 | 2.01 | 1.38 | 1.43 | 1.45 |

Table S8 shows the third validation step, which aimed to compare the effects of screening in the model. Mortality rate reductions were calculated and compared with simulation studies focusing on breast density (1, 30), the most recent Cochrane Review (31), and a comparative simulation study, which estimated screening effects in six simulation models of the Cancer Intervention and Surveillance Modeling Network (CISNET) group (32).

Table S8: Mortality reduction from screening, biennial interval vs no screening, and annual vs biennial interval

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Own calculation | Mandelblatt, Cronin (32) | Schousboe, Kerlikowske (1) | Vilaprinyo, Forné (30) | Gotzsche and Nielsen (31) |
| Biennial screening vs no screening, 50–69 years | 14.32% | 16.5% (15–23%)\* | 15.0% | 14.4% | 15.0% |
| Biennial screening vs no screening, 40–69 years | 15.74% | 19.5%(16–29%)\* |  |  |  |
| Biennial screening vs no screening, 50–79 years | 28.18% | 25%(24–30%)\* | 23.4% |  |  |
| Annual vs no screening, 50–69 years | 15.6% | 20.5%(20–33%)\* |  |  |  |
| Annual vs no screening, 40–69 years | 17.12% | 23.5% (20–43%)\* |  |  |  |
| \* Mandelblatt, Cronin (32) compare six models from the CISNET group. The values here are the median values, with the ranges in brackets. |

From Table S8, we see that the screening strategies replicate the original model’s prediction reliably. In comparison with the CISNET group’s models, our predicted effects of biennial screening are at the lower range. Annual screening has lower mortality reductions in comparison with the six CISNET models, probably because the original breast density risk estimates from the BCSC (27) had a relatively small sample size for annual screenings. The CISNET models did not include breast density, so some variation might be expected.

After having checked that the screening strategies have the expected results on mortality rates, the fourth step in the validation focused on the costing and utility assumption. As explained earlier, this model is based on the work of Schousboe, Kerlikowske (1), and thus the aim is to replicate their results. Schousboe, Kerlikowske (1) compared their results with studies by Ahern and Shen (33) and Stout, Rosenberg (34), and showed that their model closely replicates the earlier findings. We followed this example and compared our model with the earlier results to check whether the screening strategies lead to the expected cost and utility effects. In order to create a base for comparison, all incremental cost-effectiveness ratios (ICERs) were adjusted to 2008 US$, which was the base year in Schousboe, Kerlikowske (1). We used the medical care services component of the consumer price index to adjust for inflation (35). One newer study from a Canadian estimation was also used for the comparison of annual vs biennial strategies. OECD consumer price indices were used to adjust for purchasing power parities (36).

Table S9: ICER comparison between different studies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Screening strategies | Model time horizon and screening ages | Own calculation | Schousboe, Kerlikowske (1) | Ahern and Shen (33) | Stout, Rosenberg (34) | Pataky, Ismail (37) |
| Biennial vs no screening | Model from 20 to 100 years, screening from 40 to 80 years | $55,226 /QALY | $51,110 /QALY | $42,543 /QALY | $68,073 /QALY |  |
| Model from 20 to 100 years, screening from 45 to 75 years | $29,438 /QALY | $39,146 /QALY |  | $49,244 /QALY |  |
| Annual vs biennial | Model from 50 to 100 years, screening from 50 to 79 years | $380,309 /QALY |  |  |  | $728,109 /QALY |

# Sensitivity analysis

## Ranges and distributions

Table S10: Parameter ranges for univariate and probabilistic sensitivity analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | Distribution | Standard deviation | Minimum | Maximum |
| Breast cancer treatment cost | Gamma | 10% of mean | 90% of mean | 110% of mean |
| Cost of mammography | n/a | n/a | $100 | $176 |
| Disutility breast cancer | Normal | 10% of mean | 90% of mean | 110% of mean |
| Disutility false-positive result | Uniform | n/a | 0 | 0.013 |
| Invasive cancer incidence | Normal | 10% of mean | 90% of mean | 110% of mean |
| Invasive cancer mortality | Normal | 10% of mean | 90% of mean | 110% of mean |
| DCIS incidence | Uniform | n/a | 10% of mean | 110% of mean |
| Overdiagnosis | Uniform | n/a | 0% | 10% |
| Adherence probability | Uniform | n/a | 90% of mean | 135% of mean |

Schousboe, Kerlikowske (1) described the uncertainty for the univariate sensitivity analysis used in their model. They explicitly allowed 10% overdiagnosis, reflecting a potential over-identification of local cancers resulting from screening. If overdiagnosis is assumed, 10% of local cancers would not be diagnosed if screening was not conducted. Thus, local cancers have to be reduced by 10% and, correspondingly, the stage distribution has to be adjusted so that the relation between local, regional, and distant cancers remains the same.

Schousboe, Kerlikowske (1) also explicitly mention the uncertainty in the disutility of false-positive results. However, while they calculated their base case without utility reductions due to false-positive screening and allowed a decrement in the sensitivity analysis, newer models usually incorporate false-positive QALY losses. In our model, we thus use the parameter calculated by Schousboe, Kerlikowske (1) as the base case and reflect the uncertainty in the sensitivity by going back to their original assumption.

As described earlier, the sensitivity of the adherence scenarios is also tested by varying the average adherence rate around the mean value. The adherence levels were randomly changed to reflect ranges from different studies, as reported by Jacklyn, Glasziou (20). They reported adherence to breast cancer screening as reported in eight different trials with ranges from 65% to 100%. In order to reflect these in the average participation, the adherence probabilities are varied between 90 and 135% around the mean value, which transfers to probabilities between 65% and 97% (at least the slight chance of non-adherence was allowed).

In addition to these parameters, we also reflect uncertainty in other parameters, such as the cost of breast cancer treatment, the cost of mammography, the disutility of breast cancer treatment, the incidence of invasive and in situ carcinoma, and breast cancer mortality. All these parameters are varied by 10% in the univariate sensitivity analyses. In the probabilistic sensitivity analysis, we use gamma distributions for cost parameters, normal distributions for utility, incidence, and mortality, each with standard deviation of 10% around the mean value, and uniform distributions for the remaining parameters within the ranges described above.

## Univariate sensitivity

Figure S1 shows the univariate sensitivity of the screening performance indicators: days alive, days in perfect health, overdiagnosis, and false-positive results. In order to compare all variations across these four dimensions, the variations are standardized into proportional variations, measured as a percentage.

Figure S1: Tornado diagrams for variation of four screening strategies in four performance dimensions

|  |  |
| --- | --- |
|  |  |
|  |  |

# Additional results

## Significance tests

We used Welch two-sample t-tests to test significant differences in results between the adherence assumptions and to test the differences between routine and personalized screening in the univariate variation in adherence levels seen in Figure 3.

Table S11: Result of paired t-tests between adherence scenarios in two outcomes

|  |  |  |  |
| --- | --- | --- | --- |
| Screening strategy | Comparison of adherence scenarios | Cost in USD | Days in perfect health |
| SK | Pos vs uni | 0.0609 | 0.3548 |
| Neg vs uni | 0.0444 | 0.6447 |
| Curv vs uni | 0.5164 | 0.3349 |
| Pos vs neg | 0.0001 | 0.1657 |
| Curv vs pos | 0.2219 | 0.9685 |
| Curv vs neg | 0.008 | 0.1543 |
| VF | Pos vs uni | 0.1036 | 0.5272 |
| Neg vs uni | 0.0965 | 0.5959 |
| Curv vs uni | 0.259 | 0.5379 |
| Pos vs neg | 0.001 | 0.2451 |
| Curv vs pos | 0.6177 | 0.9874 |
| Curv vs neg | 0.0053 | 0.2518 |
| TDK | Pos vs uni | 0.1579 | 0.3261 |
| Neg vs uni | 0.1045 | 0.5775 |
| Curv vs uni | 0.4472 | 0.3392 |
| Pos vs neg | 0.0024 | 0.1239 |
| Curv vs pos | 0.515 | 0.9792 |
| Curv vs neg | 0.0173 | 0.1305 |
| Routine | Pos vs uni | 0.0048 | 0.4346 |
| Neg vs uni | 0.0047 | 0.7643 |
| Curv vs uni | 0.1772 | 0.444 |
| Pos vs neg | 1.626e-08 | 0.2799 |
| Curv vs pos | 0.1416 | 0.9878 |
| Curv vs neg | 3.025e-05 | 0.2871 |

Table S12: Welch one-sided two-sample significance testing, stratified screening vs routine screening

|  |  |
| --- | --- |
| Suggested strategy | Adherence level |
| 0.60 | 0.72 | 0.80 | 0.90 | 1.00 |
| SK (1) | Uniform adh. | 0.14947 | 0.10647 | 0.11084 | 0.08280 | 0.08782 |
| Positive adh. | 0.19835 | 0.13212 | 0.13043 | 0.08429 | 0.08782 |
| Negative adh. | 0.11762 | 0.11713 | 0.07769 | 0.10375 | 0.08782 |
| Curvilinear adh. | 0.17219 | 0.13287 | 0.136768 | 0.092632 | 0.08782 |
| VF (30) | Uniform adh . | 0.00731 | 0.00110 | 0.001200 | 0.000252 | 0.00010 |
| Positive adh. | 0.00873 | 0.00137 | 0.002103 | 0.000176 | 0.00010 |
| Negative adh. | 0.00531 | 0.00148 | 0.000808 | 0.000241 | 0.00010 |
| Curvilinear adh. | 0.00728 | 0.00139 | 0.001538 | 0.000278 | 0.00010 |
| TDK (38) | Uniform adh . | 0.16924 | 0.07414 | 0.081435 | 0.028201 | 0.025401 |
| Positive adh. | 0.18554 | 0.09320 | 0.094826 | 0.029594 | 0.025401 |
| Negative adh. | 0.12258 | 0.07600 | 0.053100 | 0.039162 | 0.025401 |
| Curvilinear adh. | 0.17913 | 0.09567 | 0.089740 | 0.036897 | 0.025401 |

## Subgroup analysis

The following three tables, S13–S15, present performance indicators for each of the three personalized screening approaches, stratified by the risk categories used in each strategy. Note that the risk categories are not the same, but each strategy uses its own risk definition for “low”, “average”, or “moderate” risk. Table S13 presents results for risk category “low”, for which each strategy suggests triennial screening. Differences between the strategies thus derive from the risk categorization.

Table S13: Screening performance in low risk population

|  |  |
| --- | --- |
| Suggested strategy vs no screening in risk category “low” | Units per woman |
| Incr. cost in USD | Incr. days alive | Incr. days in perfect health |
| SK (1) | Full adh. | 458.75 | 14.78 | 7.17 |
| Uniform adh . | 329.01 | 10.51 | 5.23 |
| Positive adh. | 303.96 | 9.88 | 4.96 |
| Negative adh. | 351.52 | 11.18 | 5.51 |
| Curvilinear adh. | 310.27 | 10.10 | 5.04 |
| TDK (38) | Full adh. | 1,361.52 | 17.20 | 9.17 |
| Uniform adh . | 1,258.93 | 12.31 | 6.67 |
| Positive adh. | 1,251.21 | 11.87 | 6.47 |
| Negative adh. | 1,267.76 | 12.76 | 6.88 |
| Curvilinear adh. | 1,254.78 | 12.16 | 6.61 |
| VF (30) | Full adh. | 1,740.08 | 14.47 | 6.84 |
| Uniform adh . | 1,691.54 | 10.54 | 5.02 |
| Positive adh. | 1,683.24 | 9.92 | 4.75 |
| Negative adh. | 1,699.96 | 11.09 | 5.26 |
| Curvilinear adh. | 1,685.79 | 10.12 | 4.82 |

Table S14 presents results in a similar fashion for the average risk category, which is usually defined as the population average in each study. Two strategies suggest biennial screening, but Vilaprinyo, Forné (30) suggests triennial screening.

Table S14: Screening performance in average risk population

|  |  |
| --- | --- |
| Suggested strategy vs no screening in risk category “average” | Units per woman |
| Incr. cost in USD | Incr. days alive | Incr. days in perfect health |
| SK (1) | Full adh. | 461.21 | 28.18 | 14.26 |
| Uniform adh . | 333.20 | 20.57 | 10.31 |
| Positive adh. | 332.28 | 20.86 | 10.40 |
| Negative adh. | 335.60 | 20.65 | 10.32 |
| Curvilinear adh. | 337.87 | 21.13 | 10.56 |
| TDK (38) | Full adh. | –418.06 | 30.70 | 15.17 |
| Uniform adh . | –520.65 | 22.02 | 10.81 |
| Positive adh. | –528.37 | 22.82 | 11.19 |
| Negative adh. | –511.82 | 21.63 | 10.55 |
| Curvilinear adh. | –524.80 | 22.84 | 11.25 |
| VF (30) | Full adh. | –660.49 | 26.72 | 14.18 |
| Uniform adh . | –709.04 | 19.37 | 10.29 |
| Positive adh. | –717.33 | 19.91 | 10.56 |
| Negative adh. | –700.61 | 18.99 | 10.06 |
| Curvilinear adh. | –714.78 | 19.81 | 10.56 |

Finally, Table S15 presents the performance indicators for the high-risk population, which is defined as higher than average familial risk. This risk category is still below the commonly used “high risk” category, which is reserved for genetic risk factors.

 *Table S15: Screening performance in moderate risk population*

|  |  |
| --- | --- |
| Suggested strategy vs no screening in risk category “high” | Units per woman |
| Incr. cost in USD | Incr. days alive | Incr. days in perfect health |
| SK (1) | Full adh. | –30.33 | 42.34 | 22.20 |
| Uniform adh . | –25.09 | 29.06 | 15.82 |
| Positive adh. | –29.98 | 31.18 | 16.87 |
| Negative adh. | –15.05 | 27.32 | 14.86 |
| Curvilinear adh. | –19.59 | 30.08 | 16.34 |
| TDK (38) | Full adh. | –4,321.11 | 62.42 | 31.34 |
| Uniform adh . | –4,423.70 | 42.02 | 23.21 |
| Positive adh. | –4,431.42 | 48.01 | 25.39 |
| Negative adh. | –4,414.87 | 37.01 | 21.08 |
| Curvilinear adh. | –4,427.85 | 40.86 | 22.53 |
| VF (30) | Full adh. | –6,533.42 | 75.44 | 41.53 |
| Uniform adh . | –6,581.96 | 48.92 | 29.49 |
| Positive adh. | –6,590.26 | 58.36 | 33.39 |
| Negative adh. | –6,573.54 | 36.34 | 24.75 |
| Curvilinear adh. | –6,587.71 | 30.92 | 22.53 |

## Full incremental analysis

Table S16 shows the full spectrum of the results from the probabilistic sensitivity analysis (PSA) with all possible one-on-one comparisons of the strategies under all five adherence scenarios.

Table S16: Full incremental analysis. ICERs (USD / QALY) with confidence intervals

|  |
| --- |
| Full adherence |
|  | No screening | Routine  | SK (1) | TDK (38) | VF (30) |
| No screening |  | 13,959 (13,640,14,278) | 11,310 (11,013,11,607) | 11,798 (11,495,12,102) | 6,490 (6,236,6,743) |
| Routine | 13,959 (13,640,14,278) |  | 83,394 (67,436,99,352) | 39,952 (38,859,41,045) | 54,545 (53,456,55,633) |
| SK (1) | 11,310 (11,013,11,607) | 83,394 (67,436,99,352) |  | –747 (–1515,20) | 50,530 (49,278,51,782) |
| TDK (38) | 11,798 (11,495,12,102) | 39,952 (38,859,41,045) | –747(–1,515,20) |  | 79,246 (76,479,82,012) |
| VF (30) | 6,490 (6,236,6,743) | 54,545 (53,456,55,633) | 50,530 (49,278,51,782) | 79,246 (76,479,82,012) |  |
| Risk-independent adherence |
|  | No screening | Routine  | SK (1) | TDK (38) | VF (30) |
| No screening |  | 13,061 (12,744,13,378) | 10,423 (10,128,10,718) | 10,851 (10,551,11,151) | 5,651 (5,398,5,903) |
| Routine | 13,061 (12,744,13,378) |  | 77,431 (62,652,92,210) | 45,272 (41,980,48,565) | 57,354 (55,504,59,204) |
| SK (1) | 10,423 (10,128,10,718) | 77,431 (62,652,92,210) |  | –256(–2,403,1,891) | 39,059 (8,527,69,591) |
| TDK (38) | 10,851 (10,551,11,151) | 45,272 (41,980,48,565) | –256(–2,403,1,891) |  | 77,058 (57,792,96,324) |
| VF (30) | 5,651 (5,398,5,903) | 57,354 (55,504,59,204) | 39,059 (8,527,69,591) | 77,058 (57,792,96,324) |  |
| Positive adherence |
|  | No screening | Routine  | SK (1) | TDK (38) | VF (30) |
| No screening |  | 12,464 (12,152,12,776) | 9,993 (9,702,10,284) | 10,480 (10,183,10,776) | 5,301 (5,052,5,551) |
| Routine | 12,464 (12,152,12,776) |  | 85,123 (67,083,103,163) | 49,637 (37,022,62,252) | 56,116 (54,091,58,142) |
| SK (1) | 9,993 (9,702,10,284) | 85,123 (67,083,103,163) |  | –2,989(–5,628,–351) | 50,989 (46,880,55,098) |
| TDK (38) | 10,480 (10,183,10,776) | 49,637 (37,022,62,252) | –2,989(–5,628,–351) |  | 83,624 (74,443,92,805) |
| VF (30) | 5,301 (5,052,5,551) | 56,116 (54,091,58,142) | 50,989 (46,880,55,098) | 83,624 (74,443,92,805) |  |

|  |
| --- |
| Negative adherence |
|  | No screening | Routine  | SK (1) | TDK (38) | VF (30) |
| No screening |  | 13,673 (13,350,13,995) | 10,893 (10,593,11,193) | 11,258 (10,953,11,563) | 6,008 (5,752,6,264) |
| Routine | 13,673 (13,350,13,995) |  | 116,659 (75,743,157,576) | 40,830 (30,996,50,663) | 58,475 (56,700,60,250) |
| SK (1) | 10,893 (10,593,11,193) | 116,659 (75,743,157,576) |  | 2,485(–833,5,803) | 65,483 (53,434,77,533) |
| TDK (38) | 11,258 (10,953,11,563) | 40,830 (30,996,50,663) | 2,485(–833,5803) |  | 140,545 (71,529,209,561) |
| VF (30) | 6,008 (5,752,6,264) | 58,475 (56,700,60,250) | 65,483 (53,434,77,533) | 14,0545 (71,529,209,561) |  |
| Curvilinear adherence |
|  | No screening | Routine  | SK (1) | TDK (38) | VF (30) |
| No screening |  | 12,775 (12,460,13,089) | 10,271 (9,977,10,564) | 10,658 (10,359,10,957) | 5,426 (5,175,5,676) |
| Routine | 12,775 (12,460,13,089) |  | 56,726 (–125,706,239,157) | 46,354 (42,558,50,150) | 56,186 (54,472,57,900) |
| SK (1) | 10,271 (9,977,10,564) | 56,726 (–125,706,239,157) |  | 1,110(–6,352,855) | 45,764 (32,152,59,375) |
| TDK (38) | 10,658 (10,359,10,957) | 46,354 (42,558,50,150) | 1,110(–6,352,855) |  | 84,780 (75,525,94,035) |
| VF (30) | 5,426 (5,175,5,676) | 56,186 (54,472,57,900) | 45,764 (32,152,59,375) | 84,780 (75,525,94,035) |  |

# References

1. Schousboe JT, Kerlikowske K, Loh A, et al. Personalizing Mammography by Breast Density and Other Risk Factors for Breast Cancer: Analysis of Health Benefits and Cost-Effectiveness. Annals of internal medicine. 2011; 155: 10-20.

2. Kerlikowske K, Zhu W, Hubbard RA, et al. Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. JAMA Intern Med. 2013; 173: 807-16.

3. Gierisch JM, Earp JA, Brewer NT, et al. Longitudinal predictors of nonadherence to maintenance of mammography. Cancer Epidemiol Biomarkers Prev. 2010; 19: 1103-11.

4. Tang TS, Patterson SK, Roubidoux MA, et al. Women's mammography experience and its impact on screening adherence. Psychooncology. 2009; 18: 727-34.

5. McCaul KD, Branstetter AD, Schroeder DM, et al. What is the relationship between breast cancer risk and mammography screening? A meta-analytic review. Health Psychology. 1996; 15: 423.

6. McCaul KD, Tulloch HE. Cancer screening decisions. Journal of the National Cancer Institute Monographs. 1999: 52-8.

7. Katapodi MC, Lee KA, Facione NC, et al. Predictors of perceived breast cancer risk and the relation between perceived risk and breast cancer screening: a meta-analytic review. Prev Med. 2004; 38: 388-402.

8. Hay JL, McCaul KD, Magnan RE. Does worry about breast cancer predict screening behaviors? A meta-analysis of the prospective evidence. Preventive Medicine. 2006; 42: 401-08.

9. Walker MJ, Chiarelli AM, Knight JA, et al. Perceived risk and adherence to breast cancer screening guidelines among women with a familial history of breast cancer: a review of the literature. Breast. 2013; 22: 395-404.

10. Kash KM, Holland JC, Halper MS, et al. Psychological distress and surveillance behaviors of women with a family history of breast cancer. J Natl Cancer Inst. 1992; 84: 24-30.

11. Schwartz MD, Taylor KL, Willard KS, et al. Distress, personality, and mammography utilization among women with a family history of breast cancer. Health Psychology. 1999; 18: 327-32.

12. Schwartz MD, Taylor KL, Willard KS. Prospective Association Between Distress and Mammography Utilization AmongWomen With a Family History of Breast Cancer. Journal of Behavioral Medicine. 2003; 26: 105-17.

13. Castello A, Prieto L, Ederra M, et al. Association between the Adherence to the International Guidelines for Cancer Prevention and Mammographic Density. PLoS ONE. 2015; 10: e0132684.

14. Lerman C, Caputo C, Brody D. Factors associated with inadequate cervical cancer screening among lower income primary care patients. J Am Board Fam Pract. 1990; 3: 151-6.

15. Hailey BJ. Family history of breast cancer and screening behavior: an inverted U-shaped curve? Med Hypotheses. 1991; 36: 397-403.

16. Andersen MR, Smith R, Meischke H, et al. Breast cancer worry and mammography use by women with and without a family history in a population-based sample. Cancer Epidemiology Biomarkers & Prevention. 2003; 12: 314-20.

17. Lemon SC, Zapka JG, Clemow L, et al. Mammography screening after breast cancer diagnosis in a first degree female relative: age group differences (United States). Cancer causes & control : CCC. 2006; 17: 1053-65.

18. Tice JA, Cummings SR, Smith-Bindman R, et al. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Annals of internal medicine. 2008; 148: 337-47.

19. Centers for Disease Control and Prevention. Cancer screening - United States, 2010. MMWR Morb Mortal Wkly Rep. 2012; 61: 41-45.

20. Jacklyn G, Glasziou P, Macaskill P, et al. Meta-analysis of breast cancer mortality benefit and overdiagnosis adjusted for adherence: improving information on the effects of attending screening mammography. Br J Cancer. 2016; 114: 1269-76.

21. Lidgren M, Wilking N, Jönsson B, et al. Health related quality of life in different states of breast cancer. Qual Life Res. 2007; 16: 1073-81.

22. Tosteson AN, Stout NK, Fryback DG, et al. Cost-effectiveness of digital mammography breast cancer screening. Annals of Internal Medicine. 2008; 148: 1-10.

23. Yabroff KR, Lamont EB, Mariotto A, et al. Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst. 2008; 100: 630-41.

24. Taplin SH, Barlow W, Urban N, et al. Stage, Age, Comorbidity, and Direct Costs of Colon, Prostate, and Breast-Cancer Care. J Natl Cancer Inst. 1995; 87: 417-26.

25. Vemer P, Corro Ramos I, van Voorn GA, et al. AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users. PharmacoEconomics. 2016; 34: 349-61.

26. Surveillance E, and End Results (SEER) Program. SEER\*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2015 Sub (1973-2013 varying) - Linked To County Attributes - Total U.S., 1969-2014 Counties, National Cancer Institute, DCCPS. In: Surveillance Research Program SSB, ed., 2015.

27. BCSC. Breast Cancer Surveillance Consortium (BCSC). In: BCSC, ed., Risk Factors Dataset, 2016.

28. TreeAge Software I. TreeAge Pro Healthcare. 16.2.1.0-v20160817 ed. One Bank Street Williamstown, MA, 01267 USA, 2016.

29. Microsoft. Excel 2016. Redmond, Washington, USA, 2016.

30. Vilaprinyo E, Forné C, Carles M, et al. Cost-Effectiveness and Harm-Benefit Analyses of Risk-Based Screening Strategies for Breast Cancer. PLoS ONE. 2014; 9: e86858.

31. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database of Systematic Reviews. 2011: 1-51.

32. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. Annals of internal medicine. 2009; 151: 738-47.

33. Ahern CH, Shen Y. Cost-effectiveness analysis of mammography and clinical breast examination strategies: a comparison with current guidelines. Cancer Epidemiol Biomarkers Prev. 2009; 18: 718-25.

34. Stout NK, Rosenberg MA, Trentham-Dietz A, et al. Retrospective cost-effectiveness analysis of screening mammography. J Natl Cancer Inst. 2006; 98: 774-82.

35. U.S. Bureau of Labor Statistics. Consumer Price Index. PSB Suite 3130, 2 Massachusetts Avenue, NE Washington, DC: Division of Consumer Prices and Price Indexes, 2016.

36. OECD. Prices and Purchasing Power Parities. Paris: OECD.Stat, 2016.

37. Pataky R, Ismail Z, Coldman AJ, et al. Cost-effectiveness of annual versus biennial screening mammography for women with high mammographic breast density. J Med Screen. 2014; 21: 180-8.

38. Trentham-Dietz A, Kerlikowske K, Stout NK, et al. Tailoring Breast Cancer Screening Intervals by Breast Density and Risk for Women Aged 50 Years or Older: Collaborative Modeling of Screening Outcomes. Ann Intern Med. 2016; 165: 700-12.