ARTICLE IN PRESS

VALUE IN HEALTH ∎ (2018) ∎∎∎–∎∎∎

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Personalized Mammography Screening and Screening Adherence—A Simulation and Economic Evaluation

Matthias Arnold, MBR, MSc, Dipl-Volkws[1,2,*], *Anne S. Quante, MD/MPH*[3,4,5]

[1]*Munich Center of Health Sciences, Ludwig-Maximilians-Universität, Munich, Germany;* [2]*Institute of Health Economics and Health Care Management, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany;* [3]*Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany;* [4]*Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany;* [5]*Department of Gynecology and Obstetrics, University Hospital rechts der Isar, Technical University Munich, Munich, Germany*

ABSTRACT

Objective: Personalized breast cancer screening has so far been economically evaluated under the assumption of full screening adherence. This is the first study to evaluate the effects of nonadherence on the evaluation and selection of personalized screening strategies. **Methods:** Different adherence scenarios were established on the basis of findings from the literature. A Markov microsimulation model was adapted to evaluate the effects of these adherence scenarios on three different personalized strategies. **Results:** First, three adherence scenarios describing the relationship between risk and adherence were identified: 1) a positive association between risk and screening adherence, 2) a negative association, or 3) a curvilinear relationship. Second, these three adherence scenarios were evaluated in three personalized strategies. Our results show that it is more the absolute adherence rate than the nature of the risk-adherence relationship that is important to determine which strategy is the most cost-effective. Furthermore, probabilistic sensitivity analyses showed that there are risk-stratified screening strategies that are more cost-effective than routine screening if the willingness-to-pay threshold for screening is below US $60,000. **Conclusions:** Our results show that "nonadherence" affects the relative performance of screening strategies. Thus, it is necessary to include the true adherence level to evaluate personalized screening strategies and to select the best strategy.

Keywords: adherence, breast cancer screening, decision analysis, economic evaluation, mammography, Markov model, personalized medicine.

Introduction

Many countries worldwide have introduced systematic population-based mammography screening programs. However, it remains controversial whether the benefit of screening, in terms of reduced mortality, outweighs the harm caused by overdiagnosis, referring to cancers detected at screening that would not have been detected during the woman's lifetime, as well as unnecessary diagnostic procedures involving radiation [1–4]. The Cochrane Review concluded that for every 2000 women invited for screening over a period of 10 years, 1 will be saved from cancer-related death but 10 will be treated unnecessarily, and more than 200 will suffer distress from false-positive findings [1]. The Swiss Medical Board's report 2014 concluded that "no new systematic mammography screening programs be introduced and that a time limit be placed on existing programs" [5]. With increasing knowledge about the development of breast cancer and its potential drivers, the identification of high-risk women has become more and more feasible and allows risk-based screening recommendations. It has been shown that better understanding about the individual risk of breast cancer strengthens informed choices and may thus motivate those with a higher risk to use screening opportunities [6] while reducing false-positive findings in individuals at a lower risk. A risk-based approach would therefore allocate expensive screening resources to those who would benefit the most.

Participation in breast cancer screening programs is low, especially in European countries (average 53.5%) [7]. These levels therefore do not reach the European Union benchmark of acceptable participation (>70%) for effectiveness in the reduction of mortality [8]. There is scientific evidence that screening adherence is influenced by a woman's perceived risk [9–11]. All this raises the imperative to rethink current, one-size-fits-all mammographic screening programs. It has been suggested to guide screening decisions by patients' individual risk profiles and preference [12].

Decision analytical modeling is a very useful tool to balance the benefits and harms of personalized screening under various circumstances [13–17]. However, these simulation models have not so far incorporated adherence into the decision analysis. We decided to base our simulations on a validated Markov state transition model [13], which allows the integration of

* *Address correspondence to*: Matthias Arnold, Ludwig-Maximilians-Universität München, Institut für Gesundheitsökonomie und Management im Gesundheitswesen, Ludwigstr. 28 RG, 5. OG., 80539 Munich, Germany.

E-mail: arnold@bwl.lmu.de.

1098-3015$36.00 – see front matter Copyright © 2018, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

https://doi.org/10.1016/j.jval.2017.12.022

nonadherence. This is the first study to incorporate screening adherence into the economic evaluation of personalized mammography screening, using three different risk-adherence associations.

Methods

Model Structure and Adaptation

We use a Markov state transition model of individual women, as described by Schousboe et al. [13]. The original model is validated [13] and provides an elaborate technical report, allowing for reconstruction. The Markov model assumes that healthy women may develop invasive breast cancer, ductal carcinoma in situ, or die from other causes. For women who develop breast cancer, the time spent in a healthy state before death from breast cancer or from other causes is determined depending on the cancer stage at diagnosis (local, regional, or distant). Women diagnosed with ductal carcinoma in situ can progress to invasive cancer. Figure 1 shows the state transition paths via the health states. Additional descriptions are given in Supplemental Materials found at https://doi.org/10.1016/j.jval.2017.12.022.

We use a microsimulation approach to simulate individual women with combinations of three independent risk factors—history of biopsy (28.2% of women), history of breast cancer in first-degree relative (16.1% of women), and breast density (at 50 years, 39.2% of women have heterogeneously dense and 6.4% have extremely dense tissue)—and compare three different scenario-dependent adherence behaviors (positive, negative, and curvilinear). A sample size of 3,000,000 women was found to produce robust results at relatively little variability in results within strategies compared with variability across strategies [18]. Simulations run from a start age of 50 years until the end of their life or 100 years.

Breast Cancer Incidence and Mortality

Breast cancer incidence, breast cancer mortality, and overall mortality are extracted from the original model by Schousboe et al. [13] and the Surveillance, Epidemiology, and End Results Program [19]. Schousboe et al. [13] used the Surveillance, Epidemiology, and End Results Program register data to calculate invasive and in situ breast cancer incidence rates, breast cancer mortality, and overall mortality. As the description in Schousboe et al. [13] does not provide the complete set of age-specific mortality rates, data were extracted directly from the Surveillance, Epidemiology, and End Results Program program using the updated relative survival rates from November 2014. The calculation follows the description in the original model [13].

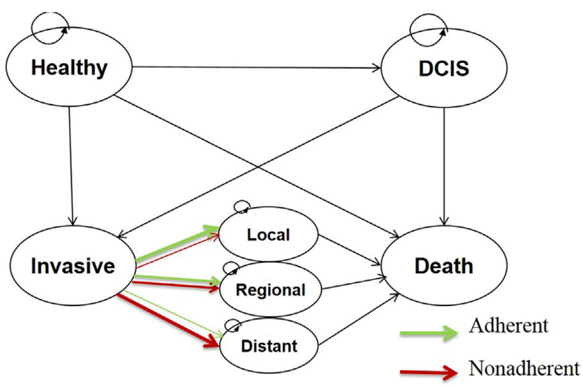

Fig. 1 – State transition model. DCIS, ductal carcinoma in situ.

Cancer incidence is stratified by the relative risk of each woman, using three risk factors: 1) breast biopsy yes/no, 2) history of breast cancer yes/no, and 3) breast density, classified by four categories 1 to 4 from the Breast Imaging Reporting and Data System [20]. Consistent with Schousboe et al. [13] and Tice et al. [21], the relative risk of invasive cancer is 1.454 or 0.938 in the presence or absence of a family history and 1.495 or 0.906 in the presence or absence of a previous biopsy. The relative risk of breast density lies between 0.388 and 1.675 depending on the Breast Imaging Reporting and Data System categorization of breast density levels and the age of the woman. We assumed that the relative risks are mutually independent and have a multiplicative effect. More details are given in Supplemental Materials.

Each woman in the simulation has a risk profile using a random combination of these three risk factors. The choice of risk factors follows the original model [13] and is derived from prevalence and relative risks from the Breast Cancer Surveillance Consortium [22]. Accordingly, 28% of all woman have a family history of breast cancer and 16% have experienced a previous biopsy. Schousboe et al. [13] assigned breast density categories independently of each other in intervals of 10 years. However, in this model, breast density is allowed to change with age, similar to Sprague et al. [16] and Trentham-Dietz et al. [15]. To reflect the natural decrease in breast density, especially at menopause, we allowed breast density to change every 10 years. With this approach, we can simulate a change in breast density and thus evaluate the complete screening strategy even when risk profile and recommendation change. We used the age-specific Breast Imaging Reporting and Data System distribution from Schousboe et al. [13] to calculate the probability of maintaining the same breast density or dropping one category every 10 years. Details can be found in the Relative risk and prevalence of breast density levels section in Supplemental Materials found https://doi.org/10.1016/j.jval.2017.12.022.

Screening Strategies

We assess mammography screening strategies for women aged between 50 and 74 years, for whom routine mammography screening is recommended. In our model, women have a combination of three risk factors reflecting a 10-year risk of breast cancer between 0.41% and 4.65%. Women with very high risk, such as the breast cancer (BRCA) susceptibility gene carriers, or high risk at younger ages have access to intensified screening including magnetic resonance imaging and are excluded from this study. Three different personalized strategies are identified from the literature with stratified screening intervals based on the combination of the three risk factors, as shown in Table 1: 1) Schousboe et al. [13], 2) Vilaprinyo et al. [14], and 3) Trentham-Dietz et al. [15]. We use the following annotation when referring to these strategies: SK, VF, and TDK.

Adherence Scenarios

From the literature, three alternative adherence scenarios were chosen. The first scenario is that women with higher perceived risk are more likely to adhere to screening. This scenario is supported by systematic reviews [23,24] and meta-analyses [9–11]. The Positive risk-dependent adherence section found in Supplemental Materials at https://doi.org/10.1016/j.jval.2017.12.022 describes the supporting evidence.

A second cluster of studies found the opposite association between perceived risk and adherence: high perceived risk may lead to psychological distress, and any form of psychological distress causes nonadherence to mammography screening. The supporting evidence consists mainly of observational

Table 1 – Personalized strategies.

Strategy	Age (y)	BI-RADS 1				BI-RADS 2				BI-RADS 3				BI-RADS 4			
		0	FH	Bio	Both	0	FH	Bio	Both	0	FH	Bio	Both	0	FH	Bio	Both
SK [13]	50–59	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2
	60–69	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2
	70–74	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2
VF [14]	50–59	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	1
	60–69	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	1
	70–74						3	3	3	3	3	3	1	3	3	3	1
TDK [15]	50–59	3	3	3	3	3	3	3	2	2	2	2	1	2	1	1	1
	60–69	3	3	3	3	3	3	3	2	2	2	2	1	2	1	1	1
	70–74	3	3	3	3	3	3	3	2	2	2	2	1	2	1	1	1

Interval length in years.

0, no additional risk factors; Bio, history of previous biopsy; BI-RADS, breast density categorization according to the Breast Imaging Reporting and Data System [20]; Both, both risk factors; *FH, family history in first-degree relative; SK, Schousboe et al.; TDK, Trentham-Dietz; VF, Vilaprinyo et al.*

studies [25–27] and experiments [28]. The Negative risk-dependent adherence section found in Supplemental Materials at https://doi.org/10.1016/j.jval.2017.12.022 describes the supporting literature.

A third group of studies exists aiming to combine the first and second hypotheses. These studies found that moderate levels of perceived risk lead to increased compliance, but low or high levels have detrimental effects. The Curvilinear risk-dependent adherence section found in Supplemental Materials at https://doi.org/10.1016/j.jval.2017.12.022 describes the supporting literature. The assumption is that moderate risk comes with moderate levels of worry or anxiety, which were found to be positively associated with screening adherence. However, very high levels of worry and anxiety produce a barrier, which was observed in women at high risk of breast cancer. The result is a curvilinear relationship between risk and adherence. This scenario is supported by observational studies [29,30] and narrative reviews [31]. As a simplification, we assume qthat the assigned risk also represents perceived risk in this simulation.

Implementation of Adherence

In this simulation study, we introduce an adherence variable that describes the likelihood of adhering to screening recommendations dependent on the risk level. We implemented nonadherence by randomly deciding for each woman whether she will attend the next screening appointment or not. If a woman does not attend one screening appointment, the distribution of cancer stages at diagnosis is similar to that in women who do not go to screening at all. The simulation continues with this distribution until the next screening invitations, in which she has a new random probability of being adherent.

The risk levels are calculated with three risk factors: breast density, family history, and breast biopsy. These risk factors can be used to calculate a risk score on the basis of the Breast Cancer Surveillance Consortium risk calculator tool developed by Tice et al. [21]. We assume that at the time point for the first screening, each woman is confronted with her risk level and is assigned a probability of adhering to screening on the basis of this risk score. We used logarithmic functions to represent the positive and negative relationships and quadratic functions to represent the curvilinear relationship. The logarithmic function for positive or negative associations was chosen to best represent the risk distribution in the population. The quadratic function for a curvilinear association was fitted to best represent the curvilinear nature reported by Andersen et al. [30]. All functions represent an effect size of 19% between risk perception and adherence, as reported by Katapodi et al. [10], and an average adherence rate of 72.4%, as reported by the Centers for Disease Control and Prevention [32]. The mathematical form can be found in the Technical implementation section in Supplemental Materials at https://doi.org/10.1016/j.jval.2017.12.022. In addition to these three risk-dependent adherence scenarios, we included two base-case scenarios: the uniform or risk-independent scenario, in which all women have the same probability of adherence (72.4%) to screening, and a full adherence scenario, in which women follow every screening invitation. Table 2 presents the three adherence scenarios, the corresponding risk levels, and the expected participation rate assuming the prevalence of risk factors as in Schousboe et al. [13].

Beneficial and Harmful Effects of Screening

The main effect of screening is to reduce mortality from breast cancer by allowing a stage shift from later cancers to early cancers at detection. With this stage shift toward earlier and more treatable cancer forms, the survival of affected women is increased. In some cases, though, mammographic screening is not sensitive enough to find all treatable cancers and, in other cases, positive results turn out to be false results. False-positive results are included using age- and breast density–dependent specificity rates, as reported by Carney et al. [33]. These false-positive screening results have both cost and utility consequences, in the form of unnecessary diagnostic workup and utility decrements of 0.013 quality-adjusted life-year (QALY), which reflects a QALY loss of 0.156 over the duration of 2 months for 50% of the women receiving false-positive results [13]. In some events, mammography screening may, however, also identify either in situ cancers, which never progress to an invasive form, or invasive cancers, which would never progress fast enough to be harmful. Once detected though, these cancers are being treated. Breast cancer screening may thus lead to overdiagnosis. *Overdiagnosis* is defined as the number of screening-detected cancer cases that would have never have been detected or treated if the woman had not been screened [15,34].

Utility and Cost

Utility values are based on the Swedish time-trade-off tariff using the five-dimensional Euroqol questionnaire, originally developed by Lidgren et al. [35] and adapted for use in an American context by Schousboe et al. [13]. The tariff uses QALY

Table 2 – Adherence probability and risk score.

Breast density (BI-RADS)	Family history (Y/N)	Previous biopsy (Y/N)	Proportion of population (%)	10-y risk of invasive breast cancer (%) [21]	Adherence probability (%)		
					Positive association	Negative association	Curvilinear association
1	0	0	2.65	0.41	62.91	81.87	65.00
1	0	1	1.04	0.62	66.14	78.65	67.44
1	1	0	0.51	0.78	67.94	76.86	69.10
2	0	0	21.26	0.85	68.61	76.19	69.77
1	1	1	0.20	1.17	71.11	73.70	72.43
2	0	1	8.35	1.28	71.81	72.99	73.18
3	0	0	28.19	1.37	72.34	72.46	73.74
2	1	0	4.08	1.60	73.56	71.25	74.92
4	0	0	8.13	1.67	73.89	70.92	75.21
3	0	1	11.07	2.05	75.49	69.32	76.22
2	1	1	1.60	2.39	76.69	68.12	76.30
4	0	1	3.19	2.50	77.05	67.77	76.16
3	1	0	5.41	2.57	77.26	67.56	76.03
4	1	0	1.56	3.13	78.80	66.02	73.82
3	1	1	2.12	3.83	80.38	64.45	68.12
4	1	1	0.61	4.65	81.90	62.93	57.29
Expected participation rate (%)					72.50	72.30	72.99

BI-RADS, Breast Imaging Reporting and Data System; N, no; Y, yes.

decrements over age and cancer stages. QALY losses are distinct in the first year and the following years to allow a differentiation between initial and follow-up treatment phases of care. To stay as close as possible to the original model, the utility weights were used as described [13]. The parameters are described in Appendix Table S2 in the Utility and cost input parameters section in Supplemental Materials found at https://doi.org/10.1016/j.jval.2017.12.022.

Similar to utility parameters, cost parameters are also extracted from Schousboe et al. [13]. Prices were inflated to represent 2016 US $ using the medical care services component of the consumer price index [36]. Screening costs represent median reimbursement rates for mammography [13]. In addition, false-positive screenings require diagnostic workup, which were calculated by Tosteson et al. [37] and also used by Schousboe et al. [13]. Treatment costs represent average costs for a patient in the respective treatment stage at diagnosis. Treatment is differentiated into initial, follow-up, and terminal treatment. Schousboe et al. [13] calculated treatment costs using estimates from Yabroff et al. [38] and Taplin et al. [39] for continuing care. The cost parameter can be found in the Utility and cost input parameters section in Supplemental Materials.

Validation

The model was validated using the AdViSHE guidance for model validation. The newly added parts of the model, especially the adherence module, were validated using the four steps described in the AdViSHE guidance: validation of the conceptual model, the input parameters, the computerized model, and the operational function. The model replicated real-world data or other published studies within a narrow range and with consistent findings. As the model was originally designed and validated by Schousboe et al. [13], aspects such as age-specific incidence, incidence ratio in the breast density categories, breast cancer mortality, and mortality reduction through screening were already assessed. As we run the model with newer data, reflecting current incidence, treatment, and mortality, we also validated the model against external data and other studies. We specifically checked lifetime incidence and mortality with different starting ages, incidence ratios for breast density, and mortality reduction for biennial and annual screening using different age intervals. Details can be found in the Validation section in Supplemental Materials at https://doi.org/10.1016/j.jval.2017.12.022.

Sensitivity Analysis

We used deterministic (univariate and multivariate) and probabilistic sensitivity analyses (PSAs) to check the impact of individual parameter uncertainty on the model and the combined effects on the overall model robustness. As we introduced the adherence variable and designed the scenarios in order to allow uncertainty about the true nature of the adherence and risk relationship, we also wanted to check the effects of the adherence scenarios and the extent of the adherence variation in sensitivity analyses. For the univariate sensitivity analysis and the PSAs, parameters are varied around 10% or the specified value ranges from the original model (see the Ranges and distributions section in Supplemental Materials found at https://doi.org/10.1016/j.jval.2017.12.022). The complete list of variables and the sensitivity ranges used are described in the Sensitivity analysis section in Supplemental Materials found at https://doi.org/10.1016/j.jval.2017.12.022. For the PSA, 1,000 runs with 100,000 trials were used. Although higher trial number reduces variance from the first-order Monte-Carlo simulation, 100,000 trials were found to produce manageable computation times (70 hours on a 26-core cluster using i5 processors) and reasonable variance increase. For the second-order Monte-Carlo simulation, we tested several iterations with 100, 300, 1000, and 3000 runs; 1000 runs were found to produce reasonably stable confidence intervals. We use cost-effectiveness acceptability curves to compare PSA results.

Results

Figure 2 describes the three personalized strategies (SK, VF, TDK), which differ in the recommended screening intervals (annual, biennial, or triennial) based on age group and a combination of three risk factors (breast density, previous biopsy, family history). For each stratum, the population share is given, which is based on the prevalence of risk factors reported by Schousboe et al. [13]. Table 3 presents screening performance indicators across the three personalized (SK, TDK, VF) and the biennial routine strategies. For each strategy, we show results for the full adherence, uniform (risk-independent) adherence, and the three risk-dependent adherence scenarios reflecting positive, negative, or curvilinear risk-adherence relationships. Compared with routine screening, SK reduces screening intervals for 26% of the population to every 3 years (Fig. 2), which leads to a slightly reduced number of screenings (Table 3). VF reduces intervals to 3 years for 97% of the population and increases the intervals to annual screening for 3% of the population. This leads to a significant reduction in the overall number of screenings. TDK maintains biennial screening for 54% of the population, reduces intervals to 3 years for 38%, and increases intervals to every year for 8% of the population (Table 3). In total, the number of screenings is reduced, similar to SK, as seen in Table 3. At the population level, all three strategies suggest fewer total screening invitations.

Uniform (Risk-Independent) Adherence versus Full Adherence

When comparing full adherence to uniform adherence, we find that nonadherence affects all performance indicators in a consistent way: a reduction of 27% to 28% across all strategies in almost all performance indicators in Table 3. As expected, screening rates drop by 28%, reflecting the average adherence of 72.4% in the uniform scenario. As a result of nonadherence, both cost and utility increments are consistently reduced by 28%. As expected, the incremental cost-effectiveness ratio is not affected because both cost and utility decrease to the same extent. If we assume full adherence, TDK and SK both outperform VF regarding days in perfect health, but both are marginally less effective than routine screening (Table 3).

Risk-Dependent Adherence versus Uniform (Risk-Independent) Adherence

For risk-associated adherence, the changes in cost and effects are not consistent. Incremental cost-effectiveness ratios decrease for risk-positive adherence (5.5% for SK, 4.5% for TDK, 7.6% for VF, and 3.3% for routine screening) and increase for risk-negative adherence (4.6% for SK, 4.5% for TDK, 6.8% for VF, and 5.8% for routine screening). All personalized strategies perform better than routine screening if we assume a risk-positive adherence:

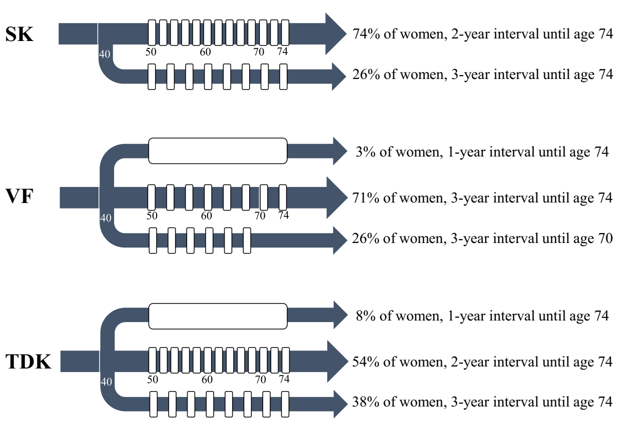

Fig. 2 – Personalized strategies, intervals, and population. SK, Schousboe et al.; TDK, Trentham-Dietz; VF, Vilaprinyo et al.

this is an expected result because we offer intensified screening to women at higher risk. Furthermore, all three personalized strategies (SK, VF, and TDK) reduce cost and increase days alive and days in perfect health significantly compared with risk-independent adherence (Table 3).

As expected, the negative risk-adherence association has exactly the opposite effect. Here, all three personalized strategies (SK, VF, and TDK) increase costs compared with risk-independent adherence, but reduce days alive and in perfect health. With curvilinear risk-adherence, costs are significantly lower for all strategies, and days in perfect health are significantly higher than with risk-independent screening. However, compared with positive risk-adherence, differences in costs and QALY are not significant (Table 3).

In a univariate sensitivity analysis (Figure 3), changing the screening adherence (in steps of 100%, 90%, 80%, 72.4%, and 60%) affects effectiveness and costs. TDK and SK produce very similar results, with only nonsignificant differences. Routine biennial screening produces the highest effect at highest cost, and VF produces significantly less effect at lowest cost. When comparing the personalized strategies, SK and TDK, to routine screening, it is important to consider the adherence level and the risk-adherence relationship. For adherence levels above 90%, SK is almost certain to produce fewer QALYs than routine screening (see the Significance tests section and Appendix Table S12 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2017.12.022). For lower adherence levels and especially positive or curvilinear relationships, the differences between SK and routine are statistically nonsignificant ($P > 0.1$). Similarly, TDK is statistically significantly less effective than routine screening only if adherence levels are above 72% (see Table S12 and the Significance tests section in Supplemental Materials). For lower adherence and especially positive or curvilinear adherence, the differences between TDK and routine screening are statistically nonsignificant ($P > 0.1$). Thus, our results show that the evaluation of personalized screening strategies compared with routine screening is dependent on the nature of the adherence level and the adherence rate.

Probabilistic Sensitivity Analysis

In the PSA, we test how important the risk-adherence relationship is when overall parameter uncertainty is allowed. Figure 4 presents the cost-effectiveness acceptability curves for all adherence scenarios, and Appendix Table S16 in Supplemental Materials (the Full incremental analysis section) found at https://doi.org/10.1016/j.jval.2017.12.022 shows the incremental cost-effectiveness ratio variation for all possible strategy combinations. The cost-effectiveness acceptability curves show the probability of a strategy being acceptable given a specific willingness-to-pay (WTP) threshold based on the net health benefit of the strategies.

If the WTP threshold is more than approximately US $5000, VF is always more acceptable than no screening (Fig. 4). From the full incremental analysis (see Appendix Table S16 and the Full incremental analysis section in Supplemental Materials), we know that the WTP threshold varies between US $5300 (positive adherence) and US $6500 (full adherence). This demonstrates that even if the WTP threshold is very low, screening, at least at very low frequencies, is more acceptable than doing nothing. VF remains the most acceptable strategy until the WTP threshold reaches almost US $50,000 (between US 49,500 for positive and US $52,500 for negative adherence). For WTP thresholds between US $50,000 and US $60,000 (US $58,500–US $60,000), SK is more acceptable than routine screening. Routine screening becomes the strategy with the highest probability of being acceptable if the WTP threshold is above US $60,000. Last, Figure 4 shows that for

Table 3 – Screening performance per adherence scenario.

Suggested strategy vs. no screening		Units per woman (95% confidence intervals)									
		Incr. total cost (US $)	Incr. cost screening + diagnostic workup (US $)	Incr. cost treatment DCIS + local (US $)	Incr. cost treatment regional + distant (US $)	Incr. days alive	Incr. days in perfect health	Proportion of overdiagnosed cancers of invasive cancers	No. of screenings	No. of false-positive results	Rejected screening invitations
SK [13]	Full adh.	371.13 (362.22–380.04)	1240.7	414.12	−1283.51	25.95 (24.98–26.93)	13.16 (12.92–13.4)	0.6 (0.58–0.63)	10.25 (10.25–10.25)	0.76 (0.76–0.76)	0
	Uniform adh.	266.69 (259.1–274.27)	898.28	300.02	−931.48	18.51 (17.68–19.34)	9.49 (9.29–9.7)	0.43 (0.41–0.45)	7.42 (7.42–7.42)	0.55 (0.55–0.55)	2.83 (2.83–2.83)
	Positive adh.	256.41 (248.79–264.03)	896.15	301.9	−941.52	18.81 (17.98–19.63)	9.63 (9.42–9.83)	0.43 (0.41–0.45)	7.39 (7.38–7.39)	0.55 (0.55–0.55)	2.87 (2.86–2.87)
	Negative adh.	277.67 (270.11–285.23)	904.16	299.8	−926.16	18.47 (17.64–19.3)	9.42 (9.22–9.63)	0.43 (0.41–0.45)	7.49 (7.48–7.49)	0.55 (0.55–0.55)	2.76 (2.76–2.77)
	Curvi. adh.	263.13 (255.5–270.75)	903.15	302.82	−942.72	18.81 (17.98–19.63)	9.63 (9.43–9.84)	0.43 (0.41–0.45)	7.45 (7.44–7.45)	0.56 (0.56–0.56)	2.8 (2.8–2.81)
VF [14]	Full adh.	181.88 (173.59–190.18)	929.38	347.47	−1094.82	22.47 (21.59–23.35)	11.64 (11.41–11.86)	0.47 (0.45–0.49)	7.64 (7.64–7.64)	0.57 (0.57–0.57)	0
	Uniform adh.	133.34 (126.3–140.38)	672.94	250.81	−790.3	16.29 (15.55–17.04)	8.46 (8.26–8.65)	0.34 (0.32–0.36)	5.53 (5.53–5.53)	0.42 (0.42–0.42)	2.11 (2.11–2.11)
	Positive adh.	125.05 (117.96–132.13)	671.66	252.73	−799.24	16.44 (15.69–17.19)	8.54 (8.35–8.74)	0.34 (0.32–0.36)	5.51 (5.51–5.52)	0.42 (0.42–0.42)	2.13 (2.13–2.13)
	Negative adh.	141.76 (134.75–148.77)	676.71	250.12	−784.96	16.21 (15.46–16.95)	8.38 (8.19–8.57)	0.34 (0.32–0.36)	5.57 (5.57–5.57)	0.42 (0.42–0.42)	2.07 (2.07–2.07)
	Curvi. adh.	127.59 (120.53–134.66)	672.17	252.58	−797.06	16.37 (15.63–17.12)	8.54 (8.35–8.73)	0.34 (0.32–0.36)	5.52 (5.52–5.53)	0.42 (0.42–0.42)	2.12 (2.11–2.12)
TDK [15]	Full adh.	371.86 (362.97–380.75)	1208.66	407.44	−1244.06	24.78 (23.82–25.75)	12.61 (12.37–12.85)	0.6 (0.58–0.63)	9.96 (9.95–9.96)	0.75 (0.75–0.75)	0
	Uniform adh.	269.27 (261.72–276.82)	875.08	294.53	−900.22	17.67 (16.85–18.49)	9.09 (8.89–9.29)	0.43 (0.41–0.45)	7.21 (7.2–7.21)	0.54 (0.54–0.54)	2.75 (2.75–2.75)
	Positive adh.	261.55 (253.95–269.15)	875.95	296.69	−910.98	18 (17.18–18.82)	9.23 (9.03–9.44)	0.43 (0.41–0.46)	7.19 (7.19–7.2)	0.55 (0.55–0.55)	2.76 (2.76–2.76)
	Negative adh.	278.1 (270.58–285.63)	877.95	294	−893.74	17.55 (16.72–18.37)	9.01 (8.8–9.21)	0.43 (0.41–0.45)	7.25 (7.25–7.25)	0.54 (0.54–0.54)	2.7 (2.7–2.71)
	Curvi. adh.	265.12 (257.53–272.71)	877.98	296.94	−909.69	17.91 (17.08–18.73)	9.23 (9.03–9.44)	0.43 (0.41–0.45)	7.22 (7.21–7.22)	0.55 (0.55–0.55)	2.74 (2.74–2.74)
Biennial routine screening	Full adh.	485.08 (476.03–494.13)	1412.23	433.39	−1360.34	28.01 (27.01–29.02)	13.81 (13.57–14.06)	0.66 (0.64–0.69)	11.69 (11.69–11.69)	0.85 (0.85–0.85)	0
	Uniform adh.	349.5 (341.79–357.2)	1022.4	314.09	−986.87	20.17 (19.31–21.02)	9.99 (9.78–10.2)	0.47 (0.45–0.49)	8.46 (8.46–8.47)	0.61 (0.61–0.61)	3.23 (3.23–3.23)
	Positive adh.	333.77 (326.03–341.5)	1012.41	315.29	−993.79	20.35 (19.49–21.21)	10.11 (9.9–10.32)	0.47 (0.45–0.5)	8.36 (8.36–8.36)	0.61 (0.61–0.61)	3.33 (3.33–3.33)
	Negative adh.	365.2 (357.51–372.9)	1036.2	314.58	−985.44	20.21 (19.35–21.07)	9.95 (9.74–10.15)	0.47 (0.45–0.5)	8.6 (8.59–8.6)	0.62 (0.62–0.62)	3.09 (3.09–3.1)
	Curvi. adh.	341.98 (334.23–349.72)	1021.6	316.43	−995.92	20.37 (19.51–21.23)	10.11 (9.9–10.32)	0.48 (0.45–0.5)	8.44 (8.44–8.44)	0.62 (0.61–0.62)	3.25 (3.25–3.25)

adh., adherence; curvi., curvilinear; DCIS, ductal carcinoma in situ; incr, incremental; SK, Schousboe et al.; TDK, Trentham-Dietz et al.; VF, Vilaprinyo et al.

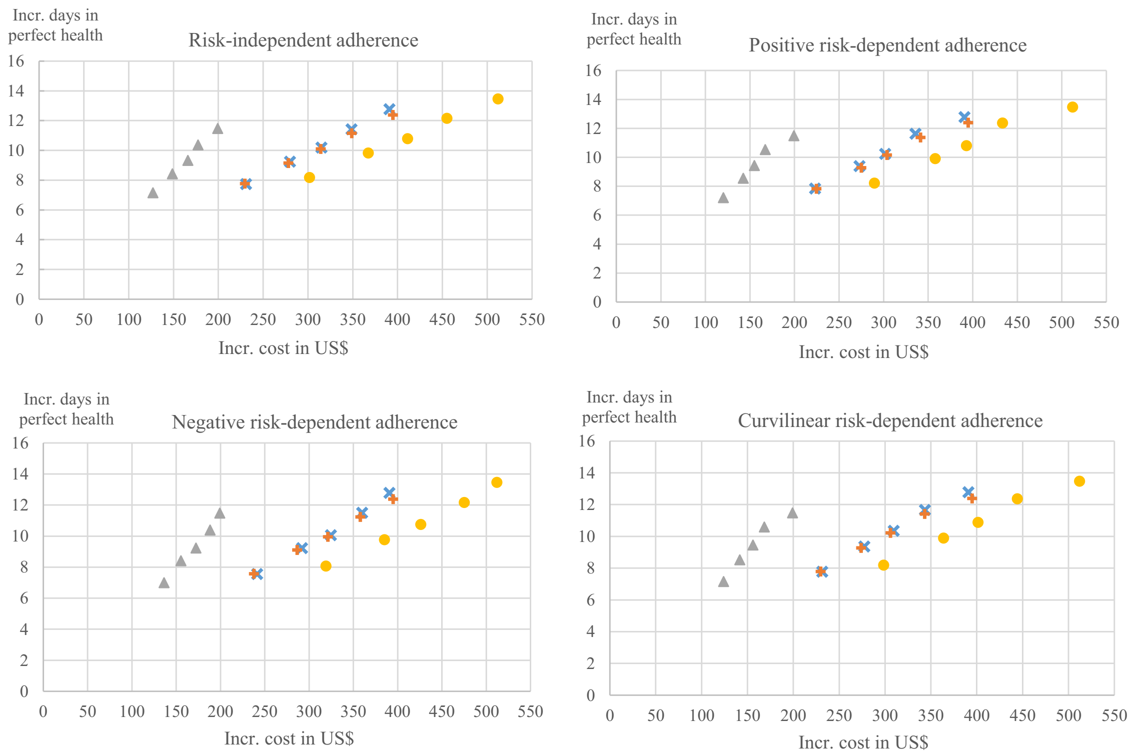

Fig. 3 – Cost-effectiveness plane, adherence variations. Incr., incremental; SK, Schousboe et al.; TDK, Trentham-Dietz; VF, Vilaprinyo et al.

WTP thresholds up to US $60,000, there are personalized strategies with higher likelihood of being acceptable than routine screening. We also see that although the WTP thresholds between the strategies change slightly, they do not change substantially. The actual character of the risk-adherence relationship is thus not as important in decision making.

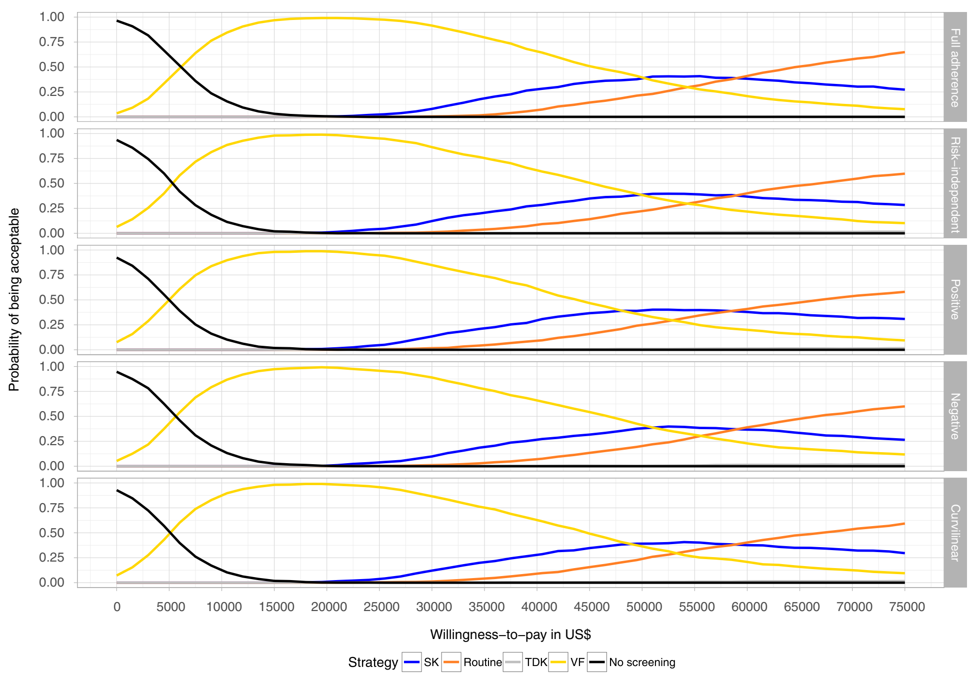

Fig. 4. – Cost-effectiveness acceptability curves.

Discussion

We used decision analytical modeling to economically evaluate personalized breast cancer screening strategies in comparison with routine mammography screening. This is the first study to incorporate the effects of nonadherence into the evaluation using three potential scenarios of risk-adherence relationship (positive, negative, or curvilinear) [13–15]. Two of the evaluated personalized strategies, SK and TDK, show similar performance (in terms of additional lifetime and quality of life) compared with routine screening, but at lower cost. The third strategy, VF, reduces screening intervals (3-year interval) substantially for most of the population, which translates into overall reduced effectiveness at substantially lower cost compared with routine screening.

Our results show that the differences in effects between SK and TDK are not statistically significant and are very close to routine screening. These results were expected, as TDK recommends annual screening for 8% of the population and triennial screening for 38% of women, whereas SK recommends triennial screening for 26%. Hence, the overall number of screenings is similar. To decide which strategy is best, we evaluated how robust the strategies are in comparison to routine screening. We showed that SK is as effective as routine screening if adherence is 80% or lower. TDK is as effective as routine screening if adherence is 60% or lower.

By definition, nonadherence reduces the number of screenings, which affects screening outcomes. Our results show that the incorporation of nonadherence into the simulation model affected the performance of the strategies. Under certain adherence conditions, personalized screening strategies may perform similarly well to routine screening, but save cost. Because all three personalized strategies were designed as cheaper alternatives to routine screening, it is not surprising that they are economically efficient alternatives for WTP thresholds below only US $60,000. However, the WTP threshold below which the VF or SK is the more effective choice compared with routine screening is dependent on the rate of adherence and also on the risk-adherence relationship. Our results show that risk-stratified screening strategies are more attractive if high-risk groups are more likely to adhere (positive adherence).

Overall, our estimations reproduce similar results to the original models in terms of beneficial and harmful screening effects, given the full adherence scenario. However, we find some differences. Vilaprinyo et al. [14] report a 20% reduction in false-positive results and overdiagnosis when comparing personalized with routine biennial screening. Our model estimates a 32% reduction in false-positive results and 28% reduction in overdiagnosis. The differences can be explained, as the personalization strategy suggested by Vilaprinyo et al. [14] starts screening for high-risk women at the age of 40 years and screening for the high- to medium-risk group at 50 years. In our study, we decided to focus on the age group above 50 years, for which we have the best evidence regarding adherence. Trentham-Dietz et al. [15] report QALY gains of 32 in 1000 from triennial screening over no screening. In a similar subgroup analysis, as reported in Supplemental Materials (see the Subgroup analysis section found at https://doi.org/10.1016/j.jval.2017.12.022), our simulation estimates QALY gains of 34 for the same comparison in a similar risk group. They also find 20% to 23% reduced false-positive screenings and 8% to 20% reduced rates of overdiagnosis from triennial screening in average risk women, a subgroup for which our simulation produces 30% reduced false-positive results and 34% reduced overdiagnosis under the assumption of full adherence. Notably, the overall overdiagnosis rate in our simulation lies between 0.3% and 0.6% of all invasive cancers, which is at the lower end of the estimation of 0% to 5% by Feig [40]. Feig [40] concluded that overdiagnosis is clinically not significant, which could be confirmed in this simulation. However, the extent of overdiagnosis is controversially discussed with estimates ranging from less than 1% to more than 30% [41].

For a decision-maker solely concerned with maximizing screening performance, the risk-adherence relationship does not affect which strategy should be chosen. If mortality reduction was the only decision-making criterion, biennial routine screening performs best, followed by SK and TDK, which reach the same number of days alive and outperform VF independently of the adherence assumptions. If avoiding false-positive results and overdiagnosis are the most important criteria, the VF strategy performs best to minimize these harmful effects. If maximizing overall quality of life (incremental days in perfect health) is the most important aim, the biennial routine performs best when full adherence is assumed. In the other adherence scenario, SK and routine screening do not show substantial differences. TDK is a noninferior alternative to routine screening only if adherence levels are lower than 80% (with a slightly higher percentage if risk is negatively associated with adherence). However, if decision makers consider cost effectiveness, the assumptions regarding adherence are very important for the decision between TDK, SK, VF, and routine screening.

The nature of the risk-adherence relationship has public health implications. From a public health perspective, a positive risk-adherence relationship is what decision makers and practitioners hope for, but it might not necessarily reflect reality. From the established research, we know that there are obstacles to screening that hinder high-risk women from attending. A positive risk-adherence relationship thus might not reflect all facets of reality. From the public health and societal perspectives, we clearly want those individuals with the greatest risk of disease to have the best access and the highest usage of preventive services. However, if research establishes that the relationship is rather curvilinear or even more negative than positive, then risk-stratified screening needs to be discussed under a new paradigm. Under these circumstances, risk-stratified screening as a tool to improve overall public health can succeed only if it includes interventions to increase screening adherence in women who would benefit the most. A sole focus on identifying cost-effective screening intervals would necessarily translate to reduce screening for high-risk women, which is not included in any recently discussed suggestion and also has very problematic implications from a public health perspective.

As with every model, some simplifying assumptions were used. Some of these assumptions limit the generalizability of our results. First, our model is based on screening performance as observed with analog film-screen mammography. We used this screening technology because the studies used to establish the risk-adherence relationships are based on film-screen mammography. This technology is still widely used, but is being replaced by full-field digital mammography. Digital mammography has the advantage of being more accurate in terms of specificity and sensitivity, and thus may improve screening performance [42], especially in women with dense breast tissue [43]. However, the performance increase in women older than 50 years is small [44], and there is no evidence that screening adherence differs between film-screen or digital screening.

Second, in the modeling of the risk-adherence relationship, we assume that the perception of risk goes hand in hand with the actual risk. However, we do know that women with a family history have a higher perceived risk [45], and thus the link between risk and perceived risk exists. However, the perception of risk is influenced by many factors, such as health literacy [46] or the method of invitation [6], which we do not incorporate into our model.

Third, we assume that the risk stratification process works perfectly in assigning women to risk clusters. We know that risk prediction has improved in recent years, but is still not perfect especially when breast density is considered [21]. We do not integrate these imperfections, which is a clear limitation of our simulation. In a simulation aiming to identify the best-stratified strategy, we strongly suggest including imperfect assignments. In analyzing whether nonadherence is important in the economic evaluation of stratified screening, we do however think that imperfect risk assignments do not have a significant effect on our analysis.

Fourth, our modeling of the risk-adherence relationship has not been validated against external data. We abstracted the three risk-adherence relationships from the literature, but we have not yet validated the functional forms against external data. In the absence of hard evidence, we used an educated guess to make the risk-adherence relationship operational. To reflect this uncertainty, we used three different forms for this relationship (positive, negative, or curvilinear) in our analysis.

Conclusions

We evaluated three personalized mammography screening strategies. One strategy, VF, produced less utility at lower cost. Two personalized strategies, SK and TDK, have been shown to provide similar performance at reduced cost compared with biennial routine screening. However, which strategy is the best depends on the level of adherence in different risk groups. We demonstrated that even small changes in adherence levels affect the performance of the screening strategy, and thus personalized screening may be an alternative to routine screening.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at https://doi.org/10.1016/j.jval.2017.12.022 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2013:CD001877.
[2] Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet 2012;380:1778–86.
[3] Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening–viewpoint of the IARC Working Group. N Engl J Med 2015;372:2353–8.
[4] Welch H, Passow HJ. Quantifying the benefits and harms of screening mammography. JAMA Intern Med 2014;174:448–54.
[5] Biller-Andorno N, Jüni P. Abolishing mammography screening programs? A view from the Swiss Medical Board. N Engl J Med 2014;370:1965–7.
[6] Edwards A, Naik G, Ahmed H, et al. Personalised risk communication for informed decision making about taking screening tests. Cochrane Database Syst Rev 2013(2):CD001865.
[7] Altobelli E, Lattanzi A. Breast cancer in European Union: an update of screening programmes as of March 2014 [review]. Int J Oncol 2014;45:1785–92.
[8] Giordano L, von Karsa L, Tomatis M, et al. Mammographic screening programmes in Europe: organization, coverage and participation. J Med Screening 2012;19:72–82.
[9] McCaul KD, Branstetter AD, Schroeder DM, et al. What is the relationship between breast cancer risk and mammography screening? A meta-analytic review. Health Psychol 1996;15:423.
[10] 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.
[11] Hay JL, McCaul KD, Magnan RE. Does worry about breast cancer predict screening behaviors? A meta-analysis of the prospective evidence. Prev Med 2006;42:401–8.
[12] Pace LE, Keating NL. A systematic assessment of benefits and risks to guide breast cancer screening decisions. JAMA 2014;311:1327–35.
[13] 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. Ann Intern Med 2011;155:10–20.
[14] 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.
[15] 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.
[16] Sprague BL, Stout NK, Schechter C, et al. Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. Ann Intern Med 2015;162:157–66.
[17] 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.
[18] Kuntz KM, Weinstein MC. Modelling in economic evaluation. In: Drummond MF, McGuire A, eds. Economic Evaluation in Health Care—Merging Theory with Practice. Oxford: Oxford University Press, 2001.
[19] Surveillance, Epidemiology, 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. Bethesda: National Cancer Institute. Surveillance Research Program SSB, 2015.
[20] Ciatto S, Houssami N, Apruzzese A, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. Breast 2005;14:269–75.
[21] 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. Ann Intern Med 2008;148:337–47.
[22] Breast Cancer Surveillance Consortium (BCSC). In: BCSC, ed., Risk Factors Dataset, 2016.
[23] McCaul KD, Tulloch HE. Cancer screening decisions. J Natl Cancer Inst Monogr 1999;1999:52–8.
[24] 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.
[25] Schwartz MD, Taylor KL, Willard KS, et al. Distress, personality, and mammography utilization among women with a family history of breast cancer. Health Psychol 1999;18:327–32.
[26] Lerman C, Schwartz M. Adherence and psychological adjustment among women at high risk for breast cancer. Breast Cancer Res Treat 1993;28:145–55.
[27] 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.
[28] Schwartz MD, Taylor KL, Willard KS. Prospective association between distress and mammography utilization among women with a family history of breast cancer. J Behav Med 2003;26:105–17.
[29] 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 2006;17:1053–65.
[30] 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 Epidemiol Biomarkers Prev 2003;12:314–20.
[31] Hailey BJ. Family history of breast cancer and screening behavior: an inverted U-shaped curve? Med Hypotheses 1991;36:397–403.
[32] Centers for Disease Control and Prevention. Cancer screening—United States, 2010. MMWR Morb Mortal Wkly Rep 2012;61:41–5.
[33] Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003;138:168–75.
[34] van Ravesteyn NT, Stout NK, Schechter CB, et al. Benefits and harms of mammography screening after age 74 years: model estimates of overdiagnosis. J Natl Cancer Inst 2015;107:djv103.

[35] 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.

[36] U.S. Bureau of Labor Statistics. Consumer Price Index. Washington, DC: Division of Consumer Prices and Price Indexes, 2016.

[37] Tosteson AN, Stout NK, Fryback DG, et al. Cost-effectiveness of digital mammography breast cancer screening. Ann Intern Med 2008;148:1–10.

[38] Yabroff KR, Lamont EB, Mariotto A, et al. Cost of care for elderly cancer patients in the United States. J Natl Cancer Instit 2008;100:630–41.

[39] 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.

[40] Feig SA. Overdiagnosis of breast cancer at screening is clinically insignificant. Acad Radiol 2015;22:961–6.

[41] Nelson H, Cantor A, Humphrey L, et al. Screening for Breast Cancer: A Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Evidence Synthesis No. 124, AHRQ Publication No 14-05201-EF-1. Rockville, MD: Agency for Healthcare Research and Quality, 2016.

[42] Hofvind S, Skaane P, Elmore JG, et al. Mammographic performance in a population-based screening program: before, during, and after the transition from screen-film to full-field digital mammography. Radiology 2014;272:52–62.

[43] Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005;353:1773–83.

[44] Vinnicombe S, Pinto Pereira SM, McCormack VA, et al. Full-field digital versus screen-film mammography: comparison within the UK breast screening program and systematic review of published data. Radiology 2009;251:347–58.

[45] Haber G, Ahmed NU, Pekovic V. Family history of cancer and its association with breast cancer risk perception and repeat mammography. Am J Public Health 2012;102:2322–9.

[46] Rutherford EJ, Kelly J, Lehane EA, et al. Health literacy and the perception of risk in a breast cancer family history clinic [published online ahead of print November 28, 2016]. Surgeon. https://doi.org/10.1016/j.surge.2016.06.003.