Mammography Screening 2.0 -
Translating Risk Adapted Screening Into Clinical Practice

HABILITATIONSSCHRIFT
Zur Erlangung der
Venia Legendi
im Fach Epidemiologie

Vorgelegt im Juli 2018

Dr. med. Anne Siobhain Quante, MPH
geboren in Neuss
Dekan der Medizinischen Fakultät:
Prof. Dr. Reinhard Hickel

Fachmentorat:
Prof. Dr. Konstantin Strauch
Prof. Dr. Rüdiger von Kries
Prof. Dr. Ortrud Steinlein
Table of Contents

LIST OF ACRONYMS 5

INTRODUCTION 6
Breast cancer – a future cancer burden 7
Breast Cancer Risk Assessment 9
Impact of modifiable risk factors on breast cancer risk 16
Economic evaluation of risk adapted screening 21
Translating risk adapted screening into clinical practice 25

DISCUSSION 28
How to calculate? 29
How to evaluate? 29
How to communicate 30

CONCLUSION 31

ABSTRACT 33

LIST OF PUBLICATIONS 34
Research articles as First or Last Author 34
Research articles as Co-Author 35
Published articles 36

CURRICULUM VITAE 37
REFERENCES 38
APPENDIX 45

Selected Original Publications 45
List of Acronyms

AAPC average annual percent change
AUC area under the curve
BC breast cancer
BCFR breast cancer family registry
BMI body mass index
BOADICEA Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm
BRCAT breast cancer risk assessment tool
CI confidence interval
CRP case risk percentile
DCIS ductal carcinoma in situ
EU European Union (EU)
GC HBOC German Consortium for hereditary breast and ovarian cancer
HD Hodgkin’s disease
HRT hormone replacement therapy
IBIS International Breast Cancer Intervention Study
KORA Cooperative Health Research in the Region of Augsburg
MRI magnetic resonance imaging
NCCN National Comprehensive Cancer Network
NICE National Institute for Health and Care Excellence
NYC New York City
25(OH)D 25-hydroxyvitamin D
OC ovarian cancer
RCT randomized controlled trial
RKI Robert Koch Institute
RLR remaining lifetime risk
ROC receiver operating characteristic
USA United States of America
Introduction

Breast cancer is the most common cancer in women (37, 52). Starting in 2005, a systematic population-based Mammography Screening Program was successfully introduced in Germany – “Mammography Screening 1.0”. Women from 50-69 years of age are currently invited for screening (2, 30). The aim is to detect breast cancer early, to accelerate access to treatment options, and ultimately reduce mortality (26, 35, 42, 73). However, there is still ongoing debate about the optimal age to initiate screening, the most efficient intervals between mammograms, and the degree of harmful effects (10, 73). Over diagnosis and exposure to radiation is an emerging concern for breast cancer screening (26, 35, 42, 73). Screening recommendations for breast cancer in the general population are currently based solely on the age and gender of an individual, despite the fact that apart from age additional genetic and non-genetic factors are known to influence cancer risk (54).

There are already target groups in Germany, for which risk-adapted screening has been implemented (40). Currently women who fulfill the inclusion criteria for the German Consortium for Hereditary Breast and Ovarian Cancer, and who carry a mutation in a known cancer gene, have access to an intensified surveillance program. The crucial point is the detection of a high risk gene (BRCA1, BRCA2, CDH1 or TP53) or a moderate gene (CHEK2, PALB2, RAD51C/D, NBN or ATM) (59). Another target group are women who were treated for Hodgkin’s disease (HD) in childhood or adolescence and who have an increased risk to develop breast cancer due to prior radiotherapy. These women also have access to an intensified surveillance program (58). Families who fulfill the inclusion criteria for the German Consortium for Hereditary Breast and Ovarian Cancer, where no mutation in any known risk gene can be identified, are more challenging. In this constellation, the decision to recommend intensified surveillance is based on the calculated risk (heterozygous risk > 20% or lifetime risk > 30%) based on Cyrillic (40).

However, for the general population only the woman’s age is taken into account for participation to the Mammography Screening Program 1.0 whereas other risk factors influence the disease risk (54). The current program therefore does not meet an individuals different needs for screening. In many women, mammography is performed without clear benefit. Other women, especially younger women, are not included in the program despite the presence of risk factors (19). Apart from familial predisposition, the time of menarche and menopause plays a part, for example, and also hormone replacement therapy and lifestyle. More recent risk models take some of these factors into account (53, 54, 67). They would allow for the implication of a
risk adapted screening approach “Mammography Screening 2.0” (18). On one hand there is a need to avoid unnecessary diagnostic procedures involving radiation in women who are unlikely to develop breast cancer, on the other hand there is a need to intensify diagnostic procedures for women at high risk. The aim of Mammography Screening 2.0 is to improve the efficiency of the screening program and to help guide screening decisions by applying individual risk profiles and preferences (14).

However, the implementation of such a strategy faces new challenges, such as the choice of the adequate prediction model, the interpretation of the results, and the ways to communicate the risks. My research focuses on the question whether we can overcome the difficulties that may arise with implementation of Mammography Screening 2.0.

Breast cancer – a future cancer burden

Breast cancer is one of the most important public health concerns, as it is worldwide the most common cancer, and the most common cause of cancer death, among women (25). The incidence of breast cancer has risen in the last decades due to changes in lifestyle, reproduction, and diet. Because of demographic ageing, it will be an even more important public health concern in the future (57).

In our research group we project the cancer incidence case number as well as the number of deaths for the most common cancer – including breast cancer – in Germany (52). For this study, cancer registry data from the Robert Koch Institute (RKI) (75) and demographic projections from the Federal Statistical Office of Germany were used (63). We projected incidences in 2020 and 2030, based on changing demographics and the change in average annual percent changes (AAPC) for the 14 most common cancer sites in women, men, and combined for both sexes (Figure 1, left). Further we also projected cancer-related deaths (Figure 1, right).
With regard to breast cancer, we could show that among women, it is projected to remain the most common malignancy with constant increase in case numbers over the next two decades (Figure 2A) (52). The good news is that the absolute number of projected cancer deaths is projected to decline over the next two decades (Figure 2A). Early detection as well as improved management has led to an improvement of prognosis of women suffering from breast cancer (1). However, breast cancer still remains the second leading cause of cancer death in women (Figure 2B) (52).

Figure 1: The formulas were used to project cancer incidences (left) and cancer-related deaths (right) (52).

Figure 2: (A) Top five projected incident cases in Germany by 2020 and 2030, among women. (B) Top five projected cancer death numbers in Germany by 2020 and 2030, among women. (52)
Breast Cancer Risk Assessment

A basic requirement for individualized screening is an accurate assessment of a woman’s cancer risk. In the United States of America (USA), annual screening mammography and magnetic resonance imaging (MRI), beginning at 30 years of age, are recommended for women with a lifetime risk of 20% or greater (61). Thus, a statistical tool is needed to estimate the probability that a currently healthy woman with specific risk factors (e.g. family history) will develop breast cancer within a certain time period (such as within ten years or in a lifetime) (22, 23).

Various statistical models have been developed for assigning absolute risks of developing breast cancer (5, 21, 47, 67). Nevertheless, the models differ in regard to the considered risk factors and how competing risks of death are being handled. More specifically, the different models can be divided into two main categories: empiric models and genetic models (55). Empiric models are based on results of epidemiological studies. Variables that are statistically significant and have a large magnitude of effect are chosen and combined using logistic regression to produce risk estimates (9). Genetic models use pedigree analysis in the form of Bayesian analysis which are based on comprehensive family history e.g. age at cancer diagnosis (9). Consequently, they can yield substantially different risk estimates. Therefore, the choice of a particular prediction model is an important aspect of individualized screening and surveillance (23).

In the two studies we have compared the performance of commonly used breast cancer risk prediction models with respect to calibration, discrimination and accuracy of very commonly used risk prediction models, focusing on different aspects:
In this study, we evaluated two commonly used risk prediction models the Breast Cancer Risk Assessment Tool (BCRAT) (20, 21) and the International Breast Cancer Intervention Study (IBIS) (67) risk assessment tool in a longitudinal New York City (NYC) cohort (n=1857) (54). The BCRAT is the most frequently used model in the USA, e.g. to determine whether a woman meets the minimum risk threshold of a five-year risk of at least 1.67% for considering tamoxifen for chemoprevention (17). The BCRAT model is an empiric model including current age, age at menarche, age at first live birth, number of previous biopsies, history of atypical hyperplasia, race/ethnicity, and number of affected first-degree relatives (15, 20, 21, 44). In contrast, the IBIS model is a genetic model including extended family history, BRCA1/2 genetic status with non genetic risk factors such as age, age at menarche, parity, age at first live birth, age at menopause, history of hormone replacement therapy (HRT) use, history of hyperplasia/atypical hyperplasia, history of lobular carcinoma in situ, height, and body mass index (BMI) (67).

It is well known that the short-term and lifetime breast cancer risks assigned to a woman by BCRA and IBIS models vary considerably. Figure 3 shows weak correlation (r=0.34) between the lifetime risks assigned by BCRAT model and IBIS model to the 1,857 participants in the current study. The IBIS model tends to assign lower risks than the IBIS model to women with a strong family history of breast cancer than does the IBIS model (54).

![Figure 3: Scatterplot of BRCAT and IBIS lifetime risks. The horizontal and vertical coordinates of points give the 1857 subjects' lifetime risks as assigned by BCRAT and IBIS, respectively. The two sets of assigned risks are only weakly correlated (Pearson correlation coefficient r=0.34) (54).](image-url)
Consequently, clinicians typically use models like BCRAT for women with average risk and models like IBIS for women with above average risk. However, the objective of this study was to compare model performance in subgroups of women typically thought to be of average risk versus subgroups classified as above average risk of breast cancer. We assessed model calibration to compare how well the model predictions agree with outcome prevalences within subgroups of the population (measurement: Hosmer Lemeshow goodness of fit statistics) (34) and we assessed the discrimination to compare its ability to discriminate those with different true risks (measurement: area under the receiver operating characteristic (ROC) curve) (29).

Figure 4 shows that overall, the agreement between assigned and observed risk was better for IBIS ($H L X^2 = 7.2$ $P$ value 0.13) than BCRAT ($H L X^2 = 22.0$ $P$ value <0.001). The IBIS model also showed better discrimination (AUC=69.5%, CI=63.8% to 75.2%) than the BRCAT model (AUC=63.2%, CI=57.6% to 68.9%) (54).

In almost all covariate specific subgroups, BCRAT mean risks were significantly lower than the observed risks, while IBIS risks showed generally good agreement with observed risks, even in the subgroups of women considered at average risk (for example, no family history of breast cancer, BRCA1/2 mutation negative).

A further useful measure of a model’s ability to discriminate for individual breast cancer cases is provided by the percentile of its assigned risk in the distribution for all non-cases, which we call its case risk percentile (CRP) (48). Figure 5 shows a scatterplot of BRCAT and IBIS models' CRPs for 83 women who developed breast cancer 10 years within 10 years of risk assignment. Points above the diagonal line
(n=46) represent cases who were better identified by the IBIS than the BCRAT model, while points below the line (n=37) were better identified by the BCRAT than the IBIS model. The mean CRP across cases for a model is its area under the curve (AUC). Using the Wilcoxon signed rank test, we also found that the median IBIS CRP was statistically significantly different than that of BCRAT model (P value 0.04).

Figure 5: Scatterplot of the case risk percentiles (CRPs). The horizontal and vertical coordinates of points give the BCRAT and IBIS CRPs, respectively, for the 83 breast cancer cases (54).

Further, the discrimination was better for the IBIS model in almost all covariate specific subgroups, except for women who had a prior biopsy where the discrimination was better for the BCRAT model.

Thus, the IBIS model, developed using extended family history and genetic data, also performs well in women considered at average risk (e.g. no family history, BRCA 1/2 mutation negative). These findings question the common clinical practice of applying risk models based on “a priori” assumptions of risks defined only by family history and genetic status (54).
Clinical guidelines for intensive surveillance including MRI screening involve estimates of remaining lifetime risk (RLR); in the USA, women with a RLR of 20% or higher meet “high risk” criteria for MRI screening. However, the clinical guidelines do not recommend which risk model to use; model predictions can differ depending on the risk factors they include and whether or not they consider the competing risk of death. In this study, we compare the risk models International Breast Cancer Intervention Study (IBIS) (67) and Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) (5, 6), which are both commonly used in clinical practice to identify women eligible for MRI screening according to the National Comprehensive Cancer Network (NCCN) guidelines (8) and have so far never been compared for USA women at high risk. Therefore using a cohort of high risk women from New York City (NYC), we compared several measures of calibration and discrimination of the IBIS and BOADICEA models (53).

The models classified different proportions of women as high risk (IBIS=59.3% vs BOADICEA=20.1%) using the RLR threshold of 20% (Table 1). Thus, if one wants to perform more MRIs, it seems that one should use the IBIS model. However, using the 10-year threshold of 3.34%, the difference was smaller (IBIS=52.9% vs BOADICEA=43.2%) (Table 1). These differences could in part be because of the higher RLR upper age bound used by the IBIS (85 years) compared with the BOADICEA (80 years) model. We found the discordance was less when we defined high risk by a 10-year risk of 3.34% of higher (which is roughly equivalent to the NCCN 5-year risk of 1.67% (8)) (53).
Table 1: Classification of 1764 study subjects into high and low risk groups by the IBIS and BOADICEA models, using remaining lifetime risk and 10-year risk thresholds (53).

<table>
<thead>
<tr>
<th></th>
<th>Remaining lifetime risk*</th>
<th>IBIS</th>
<th>BOADICEA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low†</td>
<td>High</td>
<td>Total</td>
</tr>
<tr>
<td>Low</td>
<td>704</td>
<td>705</td>
<td>1409</td>
</tr>
<tr>
<td>High</td>
<td>13</td>
<td>342</td>
<td>355</td>
</tr>
<tr>
<td>Total</td>
<td>717</td>
<td>1047</td>
<td>1764</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>10-year risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low‡</td>
</tr>
<tr>
<td>Low</td>
<td>775</td>
</tr>
<tr>
<td>High</td>
<td>55</td>
</tr>
<tr>
<td>Total</td>
<td>830</td>
</tr>
</tbody>
</table>

* Defined as risk from assessment until age 80 years (BOADICEA) or 85 years (IBIS). BOADICEA = Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; IBIS = International Breast Cancer Intervention Study.
† Remaining lifetime risks of 20% or greater.
‡ Ten-year risks of 3.94% or greater.

IBIS risks (mean=4.9%) were better calibrated to observed breast cancer incidence (5.2%, 95% confidence interval (CI) = 4.2% to 6.4%) than were those of BOADICEA (mean=3.7%) overall and with quartiles of model risk ($P$ value 0.20 by IBIS and $P$ value 0.07 by BOADICEA). Both models gave similar discrimination, with area under the curves (AUC) of 0.67 (95% CI=0.61 to 0.73) using IBIS and 0.68 (95% CI=0.62 to 0.74) using BOADICEA models (53).

To supplement the AUC based comparison of IBIS and BOADICEA models’ discrimination, we also compared the IBIS and BOADICEA models’ case risk percentiles (48) of the 79 women who developed breast cancer within 10 years of recruitment. The vertical dashed line gives the threshold corresponding to 80% specificity for the BOADICEA model, while the horizontal dashed line gives the corresponding 80% specificity threshold for the IBIS model. The Figure 6 shows that seven cases were correctly deemed high risk by the IBIS but not by the BOADICEA model, while four cases were correctly identified as high risk by the BOADICEA but not by the IBIS model. This comparison gives a sensitivity of 33 of 79, or 41.8% for the IBIS model, and of 30 of 79, or 38.0% for the BOADICEA model (53).
Our research showed that clinical guidelines based on RLR for high risk women are limited by discordance between commonly used risk models. Guidelines based on short term risks would be more useful, as models are generally developed and validated under a short time horizon (<10 years). Nevertheless, both IBIS and BOADICEA models still underestimated 10-year breast cancer risks in our cohort, with the discrepancies larger for BOADICEA than the IBIS model. The data suggest that the improved IBIS calibration reflects its inclusion of non-genetic risk factors (53).
Impact of modifiable risk factors on breast cancer risk

Breast cancer is caused by a complex interplay of many genetic and non genetic risk factors. Therefore, the choice of a particular prediction model is an important aspect of clinical counseling. Further, many modifiable risk factors, such as physical activity and alcohol intake, are not included in current risk assessment tools. Thus, integration of risk reduction strategies based on modifiable factors is limited to the modifiable factors present in a given risk model.

Body mass index (BMI) is one of the few modifiable risk factors included in the IBIS model (67). To illustrate the potential impact of risk factor modification in model based risk assessment, we evaluated the IBIS model with and without BMI, for predicting future breast cancer occurrence in a prospective cohort of 665 postmenopausal women (51). We focused on the role of BMI in postmenopausal women because of its positive association with breast cancer risk among postmenopausal, but not premenopausal women (12, 68).

For each of 665 postmenopausal women we calculated the 10-year risks as assigned by the IBIS model with and without inclusion of the BMI (Figure 7). Overall, the accuracy of the IBIS model (overall agreement between observed and assigned risks) and discrimination (AUC concordance between assigned risks and outcomes) were similar with and without the BMI. In women with BMI > 25 kg/m², adding the BMI improved discrimination (AUC=63.9% and 61.4% with and without BMI, P value >0.001). Using the commonly used 3.4% threshold for high risk status, the addition of the BMI reclassified 11 women from low to high risk, and 11 women from high to low risk (51).
Further, we also analyzed the BMI’s influence on IBIS’10 year risk for five hypothetical women aged 50 years with varying levels of hereditary risk, keeping all other factors constant. For each woman, we examined the difference between her IBIS model risk with BMI of 27 versus 21 kg/m² in terms of number of affected first-degree relatives and BRCA1 mutation carrier status. Table 2 shows that this difference increases with a woman’s hereditary risk, ranging from 0.3% for women without affected relatives or BRCA1 mutation to 4.5% for those with three affected relatives and a BRCA1 mutation. This contrast shows that a woman at high hereditary risk can move across the 10-year threshold of 3.4% used to increase screening strategies by increasing her BMI. Moreover, obese women classified as high risk (i.e. as having ≥3.4% 10-year risk) could be reclassified as low risk by changing BMI alone, although this would require large weight reduction (51).

Table 2: Effects of including BMI in IBIS model according to inherited risk factors (51).

<table>
<thead>
<tr>
<th>Inherited risk factors</th>
<th>10-year risk (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. affected first-degree relatives</td>
<td>BMI = 27 kg/m²</td>
<td>BMI = 21 kg/m²</td>
</tr>
<tr>
<td>0</td>
<td>1.8</td>
<td>1.5</td>
</tr>
<tr>
<td>1</td>
<td>3.8</td>
<td>3.0</td>
</tr>
<tr>
<td>2</td>
<td>5.0</td>
<td>4.0</td>
</tr>
<tr>
<td>3</td>
<td>9.7</td>
<td>7.8</td>
</tr>
<tr>
<td>3 + BRCA1 mutation carrier</td>
<td>24.6</td>
<td>20.1</td>
</tr>
</tbody>
</table>
More generally, this argument indicates that even when risk factors have limited impact at a population level, they can have a large impact on how an individual is classified into categories affecting their screening and chemoprevention counseling. We recommend that women be informed about both absolute and relative risk reductions when counseled for breast cancer prevention (51).

As a potentially modifiable risk factor, 25-hydroxyvitamin D [25(OH)D] has become a major topic in cancer research (16). The question is whether vitamin D is protective against cancer. To date, only a few prospective studies have examined the relationship between serum 25(OH)D concentration and total cancer risk. We designed a population-based prospective cohort study to test the association between serum 25(OH)D concentration and the development of any cancer as well as specific common cancer types. The individuals for the present study were selected from the participants of the KORA (Cooperative Health Research in the Region of Augsburg) study (33, 74). We analysed data from the KORA cohort study including 2,003 initially cancer-free participants with baseline serum 25(OH)D measurements. We identified 69 participants who developed cancer during the 7-year follow up period (5.4 cases per 1000 person year). The most common cancers were prostate (2.9 cases per 1000 person year), breast (2.5 cases per 1000 person year) and colorectal cancer (0.8 cases per 1000 person year) (11).

We used Cox proportional hazard models to assess the association between 25(OH)D levels and incident cancer risk (Table 3). Overall, we observed no significant relationship between serum 25(OH)D levels and cancer risk. The hazard ratio (HR) [95% CI] per 1 ng/ml increase in 25 (OH)D for this relationship was 1.00 [0.97-1.04] adjusting for age, sex, BMI, and season of blood draw. This was also true in subgroup analysis for prostate cancer (HR 0.95 [0.88-1.03]), breast cancer...
(HR [95% CI] 1.03 [0.97-1.09]), and colorectal cancer (HR [95% CI] 0.97 [0.88-1.07]) (11).

Table 3: Hazard ratios for risk of developing cancer according to the baseline serum 25 (OH) D levels (11).

<table>
<thead>
<tr>
<th></th>
<th>Continuous</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases/controls, n</td>
<td>69/1,934</td>
<td>21/647</td>
<td>25/634</td>
<td>23/653</td>
<td></td>
</tr>
<tr>
<td>Model 1*, HR (95% CI)</td>
<td>1.00 (0.97-1.04)</td>
<td>1.0</td>
<td>1.12 (0.62-2.03)</td>
<td>0.96 (0.51-1.81)</td>
<td>0.896</td>
</tr>
<tr>
<td>Model 2*, HR (95% CI)</td>
<td>1.00 (0.97-1.04)</td>
<td>1.0</td>
<td>1.15 (0.63-2.10)</td>
<td>1.00 (0.52-1.90)</td>
<td>0.810</td>
</tr>
<tr>
<td>Prostate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases/controls, n</td>
<td>18/941</td>
<td>7/310</td>
<td>8/308</td>
<td>3/323</td>
<td></td>
</tr>
<tr>
<td>Model 1*, HR (95% CI)</td>
<td>0.95 (0.88-1.03)</td>
<td>1.0</td>
<td>0.92 (0.33-2.63)</td>
<td>0.29 (0.07-1.17)</td>
<td>0.206</td>
</tr>
<tr>
<td>Model 2*, HR (95% CI)</td>
<td>0.96 (0.88-1.03)</td>
<td>1.0</td>
<td>0.92 (0.31-2.73)</td>
<td>0.29 (0.07-1.23)</td>
<td>0.240</td>
</tr>
<tr>
<td>Breast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases/controls, n</td>
<td>16/993</td>
<td>4/337</td>
<td>5/326</td>
<td>7/330</td>
<td></td>
</tr>
<tr>
<td>Model 1*, HR (95% CI)</td>
<td>1.03 (0.97-1.09)</td>
<td>1.0</td>
<td>1.47 (0.39-5.61)</td>
<td>2.15 (0.59-7.88)</td>
<td>0.309</td>
</tr>
<tr>
<td>Model 2 BC, HR (95% CI)</td>
<td>1.03 (0.97-1.10)</td>
<td>1.0</td>
<td>1.34 (0.35-5.18)</td>
<td>1.99 (0.53-7.55)</td>
<td>0.336</td>
</tr>
<tr>
<td>Colon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases/controls, n</td>
<td>10/1,934</td>
<td>4/647</td>
<td>2/634</td>
<td>4/653</td>
<td></td>
</tr>
<tr>
<td>Model 1*, HR (95% CI)</td>
<td>0.97 (0.88-1.07)</td>
<td>1.0</td>
<td>0.38 (0.07-2.17)</td>
<td>0.59 (0.12-2.83)</td>
<td>0.694</td>
</tr>
<tr>
<td>Model 2*, HR (95% CI)</td>
<td>0.98 (0.89-1.08)</td>
<td>1.0</td>
<td>0.38 (0.07-2.24)</td>
<td>0.64 (0.13-3.10)</td>
<td>0.710</td>
</tr>
</tbody>
</table>

| aTertile cut-off points: T1 = 2.0–11.3 ng/ml, T2 = 11.4–17.9 ng/ml, T3 = 18.0–73.3 ng/ml. |
| bTwo-sided p value, obtained by the Wald chi-square statistic, for the overall trend of 25(OH)D level as a continuous variable. p values of < 0.05 were considered statistically significant. |
| cModel 1 adjusted for age, sex, BMI, and season of blood draw. |
| dModel 2 adjusted for variables in Model 1 + physical activity (active vs. inactive), smoking status (regular smoker, occasional smoker, ex-smoker, or non-smoker), alcohol consumption (g/day), and vitamin D supplementation (yes vs. no). |
| eModel 2 BC adjusted for variables in Model 2 + menopausal status (pre- vs. postmenopausal); Note: Breast and prostate cancer models were not adjusted for sex. |

However, the mean serum 25(OH)D concentrations were well below levels recommended by the Endocrine Society (20-30 ng/ml) for both the cases and non-cases (Figure 8) (32).
Figure 8: Distribution of serum 25-hydroxyvitamin D (25(OH)D) concentration in the study population. The cut-off for defining vitamin D deficiency (20ng/ml) is indicated by the arrow.

The prevalence of 25(OH)D sufficiency was low, as described previously (56). Our study found no protective effect of 25(OH)D against developing cancer. However, studies with more participants and additional measurements of 25(OH)D are still needed accurately clarify the relationship between 25(OH)D and total cancer risk (11).
Economic evaluation of risk adapted screening

Because of the high costs involved in cancer screening and treatment, cost effectiveness is a key point in new strategies for individualized prevention. Mammography screening is the only cancer detection program in Germany that meets European Union (EU) directives entailing systematic invitation and quality assurance monitoring (2, 30). However, recent findings indicate that the cost effectiveness of current national screening programs could be improved (49). Screening procedures, such as mammography, require an expensive medical infrastructure, which entail both potential benefit and harm to participants. A more focused view on the individual risk for cancer may motivate individuals with high risk to use screening opportunities while reducing false positive findings and alleviating concerns of individuals at lower risk. A risk based approach would therefore allocate expensive screening resources to those who benefit the most from it.

Individualized breast cancer screening has so far been economically evaluated under the assumption of full screening adherence. However, participation in breast cancer screening programs is low, especially in the EU (average 53.5%) (3). These levels do not reach the EU benchmark of acceptable participation (>70%) for effectiveness in the reduction of mortality (3, 24). There is scientific evidence to support that screening adherence is influenced by a woman’s perceived risk (28, 39, 46). Decision analytical modelling is a very useful tool to balance benefits and harms of screening under a variety of circumstances. Recent studies have already taken up the challenge to weigh the balances of individualized screening (43, 60, 62, 66, 70). However, these simulation models so far have not incorporated adherence into the decision analysis. This is the first study to identify three different risk adherence associations and to incorporate screening adherence into the economic evaluation of individualized mammography screening (7). In this study we base our simulations on a validated Markov transition model (60), which allows the integration of nonadherence. Figure 9 shows the state transition via the health states. The Markov
model assumes that healthy women may develop invasive breast cancer, ductal carcinoma in situ (DCIS) or die of other causes. Depending on the cancer stage at diagnosis, local, regional or distant, the time spent in this health state before death from breast cancer or from other causes is determined (7).

Figure 9: State transition model (7).

We assess mammography screening strategies for women aged between 50 and 74 years. In our model, women have a combination of three risk factors (breast density, previous biopsy, family history).

We use a micro-simulation approach to simulate individual women with combinations of the three risk factors. Simulations run from a start age of 50, until the end of their life or 100 years.

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 relation relationship.

Further, these three adherence scenarios were evaluated in three individualized strategies which were identified from the literature with stratified screening intervals based on the combination of the three risk factors: 1) Schosboe et al (SK) (60), 2) Vilaprinyo et al (VF) (69) and 3) Trentham-Dietz et al (TDK) (66).

Figure 10 describes the three individualized 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) (7).
In a univariate sensitivity analysis (Figure 11), 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 lower cost. When comparing the individualized strategies, SK and TDK, to routine screening, it is important to consider the adherence level and the risk adherence relationship. For adherence level above 90%, SK is almost certain to produce fewer QALYs than routine screening. For lower adherence levels and especially positive or curvilinear relationships, the differences SK and routine screening are statistically nonsignificant. Similarly, TDK is statistically significantly less effective than routine screening only if adherence levels are above 72%. For lower adherence and especially positive or curvilinear adherence, the differences between TDK and routine screening are statistically nonsignificant (7).
Thus, our results show that the evaluation of individualized screening strategies compared with routine screening is dependent on the nature of the adherence level and the adherence rate (7). All three individualized strategies were designed as cheaper alternatives to routine screening. Under certain adherence conditions, individualized screening strategies may perform similarly well to routine screening, but save cost. Our results show that risk-stratified strategies are more attractive if high-risk groups are more likely to adhere (positive adherence).

In conclusion, we show that “nonadherence” affects the relative performance of screening strategies. Thus, it is necessary to include the true adherence level to evaluate individualized screening strategies to select the best strategy (7).
Translating risk adapted screening into clinical practice

The current Mammography Screening Program (Mammography Screening 1.0) has been the subject of criticism for some time (26, 35, 37). Invitation to take part is currently based on the risk factors of age and female sex, whereby women with an above average risk are screened too seldom and women with a low risk are possibly screened too often. Ultimately, the vision is to translate a risk adapted screening approach (Mammography Screening 2.0) into clinical practice by focusing screening efforts on those individuals who are most likely to develop cancer (18).

In Mammography Screening 2.0 dealing with the risk models represents an additional challenge (55). This is where our current research project RISIKOLOTSE.DE comes in. An online platform will be generated which provides information and tools that will allow the breast cancer risk to be calculated, understood and evaluated. The target groups are doctors and laypersons: doctors will be supported in risk communication and counseling, laypersons in weighing the benefits and risks of taking part. The decision-making process for participation will be facilitated thereby. As an initial assessment of the demand for this project we organized a focus group with 15 physicians and representatives of the health service. The discussion was analyzed with qualitative methods (18, 45).

As an introduction, two brief cases were presented (Figure 12). The experts were to decide whether they would advise the women for or against mammography. For the 42-year old woman in case 1, whose cousin had breast cancer, despite an increased breast cancer risk, there was no clear advice. In contrast, nearly all of them advised the 51-year old woman in case 2, with average breast cancer risk, to take part in the
mammography screening. The two case examples show were the weakness of the current mammography screening lie. The invitation is currently based on the age of women between 50 and 69 years. This screening recommendation appears to represent a binding guideline for doctors. On the other hand, there is no uniform recommendation on mammography for women under 50 years. However, individual risk calculation for the IBIS model risk calculation procedure, which includes other risk factors, results in an increased breast cancer risk for 42-year old woman and an average risk for the 51-year old compared with the general population. Nevertheless, only the 51-year old is included in the screening program according to the current standards (18).

Figure 12: Voting result of the case examples (18).

The participants admitted that when counseling about breast cancer screening they would emphasize positive aspects over negative aspects. General practitioners in particular reported that they are confronted “really often” with questions about the current mammography screening. The physicians were uncertain whether they were competent enough for counseling and were self-critical, e.g." It is often the case that the women but also the doctors don’t understand it." (gynecologist) or " “I am already a bit uncertain because I originally thought that screening can really only be good (..) But it is not quite so simple." (general practitioner).

The concept of individualized risk adapted screening was assessed positively overall by the participants. The limitation in MammographyScreening 1.0 to the risk factors age and sex was criticized unanimously: “We do know that that doesn’t suffice" (gynecologist) Thus, “mammograms are done in women who derive no benefit from them” (gynecologist). In this connection, the financial aspects for the health care system were also addressed: “Because of limited resources, it must be considered (...) whether it is actually necessary for us to screen all women" (public health service).
The participants were in agreement that only physicians should interpret the results of the risk calculation: “The (medical) interpretation is always needed” (gynaecologist). Concerns were raised, if laymen would use risk calculators by themselves: “You only need to imagine a woman with a family history (…), who keeps on clicking, forgets something and lands at a supposedly low risk and tells herself: everything’s OK ” (gynecologist). Possible positive aspects of independent use of risk calculators by lay persons (empowerment) were not mentioned by the participants.

The biggest challenge for the implementation is the lack of time and the complexity of risk communication, which would overtax many doctors. Counseling about the results of an individualised risk calculation is even more challenging: “I must classify the risk and we already see from genetics, that that is not so simple.” (gynaecologist). The objection was made that doctors could learn these competencies: “I find that every doctor must be able to handle the subject of risk communication” (gynaecologist).

The experts from the focus group approved the idea of risk adapted screening (Mammography Screening 2.0). “I think there are women who have such a low risk that they need less or no screening” (radiologist). On the other hand, women with an increased risk would benefit from earlier, more frequent, or longer participation, and from additional investigations such as ultrasonography or MR imaging. Mammography Screening 2.0 could thus lead to “provision of better care for the overall population, (…) by simply redistributing them [=resources]” (radiologist).

Overall, the suggestions and new ideas from the focus group ranged from administrative and regulatory changes to new forms of counselling and adaptable practice aids. An important indicator for the RISIKOLOTSE.DE conception and for planning future surveys was that risk calculation for Mammography Screening 2.0 was regarded as purely medical function and that the concept of participatory decision making played hardly any part in the discussion (18).
Discussion

Breast cancer is one of the most important public health concerns, as it is the most common cancer, and also the most common cause of cancer death, among women (1). The incidence of breast cancer has risen in the last decades due to changes in lifestyle, reproduction and diet. Because of demographic ageing, it will be an even more important health issue in the future (52). It has been estimated that around 30% of all breast cancers are attributable to modifiable factors (e.g. excess body weight, physical inactivity and alcohol intake) (38). Even though there is enormous potential to reduce the burden of disease in the general population by primary prevention, the majority of cancer cases cannot be avoided (76). Therefore, secondary prevention by means of screening and early treatment appears important, and there has been strong interest in implementing screening strategies that will detect early breast cancer, thereby reducing mortality (2).

Approximately Ten years after successfully introduction in Germany in 2005 of a systematic population-based Mammography Screening Program, many controversies have arisen (26, 36, 37, 72). In particular on the benefit of screening, in terms of whether the benefit of reduced breast cancer mortality, outweighs the harm caused by over diagnosis of cancers detected at screening that would not have been detected during the woman's lifetime, as well as unnecessary diagnostic procedures involving radiation (26, 36, 37, 72). The Cochrane Review concluded that, for every 2000 women invited for screening over a period of 10 years, one will be saved from cancer-related death but 10 will be over diagnosed (27). The Swiss Medical Board’s report 2014 concluded that “no new systematic Mammography Screening Program be introduced and that a time limit be placed on existing programs” (10).

All of this raises the need to re-think current mammography screening programs e.g. Mammography Screening 1.0. Is it appropriate to perform screening in every woman aged 50-69 years, or should uniform population screening be replaced or complemented with an individualized screening approach: such as Mammography Screening 2.0 (13)? In light of the recent emergence of more accurate risk prediction tools and the fact that trade-offs between risks and benefits may vary on the individual level, an individualized program is likely to represent a more effective approach than population-wide screening. However, an individualized screening strategy faces new challenges, such as (i) how to calculate, (ii) how to evaluate and (iii) how to communicate (55).
My research has addressed ways of translating risk adapted screening into clinical practice.

How to calculate?

A basic requirement for risk adapted screening is an accurate assessment of a woman’s absolute breast cancer risk. Therefore, the choice of a particular prediction model is an important aspect of individualized screening and surveillance (53, 54).

So far clinicians use different breast cancer risk models for women considered at average and above average risk, based largely on their family histories and genetic factors. However, oncologists and genetic counsellors would be well served by a single model that avoids having to choose among several models on the basis of patient characteristics (54).

To address this question, we compared different breast cancer risk assessment tools in a longitudinal NYC cohort (54). In summary, we found that the IBIS model surpassed the BCRAT model in this cohort whose risks span the continuum of breast cancer risk. This was true in subgroups containing women typically considered to be of average risk (for example, no family history of breast cancer, BRCA1/2 negative).

We could show that a risk model that has been developed based on extended family history and genetic data, such as the IBIS model, also performs well in women considered at average risk (54). Further, we could show that IBIS and BOADICEA models were both well calibrated, even though IBIS performed slightly better. Both models gave similar discrimination, with area under the curve of 0.67 using IBIS and 0.68 using BOADICEA (53).

A limitation to our study is that our validation study and other validation studies have been conducted with small numbers (4, 41, 50, 53, 54). There is a need for much larger studies in high risk populations and in the general population (23).

However, we could show that currently used risk models, such as IBIS and BOADICEA, are well calibrated and conclusively - with an AUC almost 0.7 - already provide a useful perspective for individualized screening as they perform better than considering age alone (23).

How to evaluate?

Our ultimate goal is to categorize women according to their risk of breast cancer as accurately as possible, based on their profile of genetic and non-genetic risk factors, and to recommend a more individualized screening program. However, at the
moment we do not have an international agreement how best to define the “high risk” group.

The NCCN guidelines (8) recommend that high risk women consider risk reduction strategies and annual MRI starting at the age of 30 years. The NCCN definition of high risk includes women who have a RLR of 20%. However, it is not specified which risk model to use. We found that the life time risks assigned by IBIS or BOADICEA differ substantially and thus the models classified different proportions of women as high risk: 60% when using IBIS compared to 20% using the BOADICEA. Thus, if one wants to perform more MRIs, it seems that one should use the IBIS model.

A further challenge for risk adapted screening is that clinical guidelines for breast cancer chemoprevention and MRI screening involve estimates of RLR. We found that implementing these RLR-based guidelines is problematic for two reasons (53). First, the models disagree on the definition of a woman’s RLR with the IBIS model using age 85 years and the BOADICEA model using age 80 years. We found that the RLRs assigned by the IBIS and BOADICEA models differ substantially and that these differences were reduced when comparing the models’ predicted 10-year risks. Second, it is not feasible to assess model RLR predictions, because we cannot observe breast cancer outcomes during the remaining lifetimes of cohort subjects. In fact, risk models are generally developed and validated over a fixed horizon (e.g. five or ten years of follow-up), making estimates of RLR less precise (53).

Thus, a greater recognition of the limitation of RLRs estimates is needed. Increased clinical use of shorter fixed time horizons when conveying risk, may be particularly important in order to improve the validity of risk assessment. We recommend that physicians use time periods of briefer duration, such as 5- or 10-year risks (53).

**How to communicate**

It is apparent in the literature that communication in the framework of Mammography Screening 1.0 is already a challenge. Women should be enabled to make an informed decision (64). The informed consent assumes that benefits and risks of screening are understood, correctly interpreted and applied to their own situation. A requirement for the health competence or health literacy is a basic understanding of statistics. However the reality is, that statistical statements about positive and negative effects of screening are often misinterpreted by both doctors and by laypersons seeking advice (71). The benefit is sometimes massively overestimated while the risks are rather trivialized (31). In the future Mammography Screening 2.0 dealing with the risk models may present an additional challenge. Many doctors
assume in general that they already practice joint decision-making. Studies suggest that the assessment is deceptive and that there is a perception–reality gap (65). Results from our focus group showed that the experts emphasized that risk calculation in particular should be reserved to doctors; lay persons were not trusted by the majority to be able to take responsibility for this. The biggest challenge for the implementation is the complexity of risk communication, as counseling about the results of an individualized risk calculation is more challenging (18).

Overall, the experts form the focus group approved the idea of risk adapted screening. However, there is a need to implement participatory decision-making in routine medical practice, and also a need for fundamental schemes for risk communication (18).

Conclusion

Although the current Mammography Screening 1.0 is mainly age dependent, there are already target groups in Germany, for which risk-adapted screening has been implemented (40). Currently, indisputably women who carry a mutation in a known cancer gene, have access to an intensified surveillance program (59). Access to intensified surveillance is more controversial for women where the decision is based on the calculated risk (heterozygous risk >20% or lifetime risk > 30%) based on Cyrillic (59). This risk calculation method is outdated and needs to be revised (55). However, changing the risk calculation method facilitates a big challenge for the German Consortium of Hereditary Breast and Ovarian Cancer (55). The target group which will have access to intensified surveillance will change and we need to communicate the changes in decision rules to our patients. Some women will “no longer have an increased risk” and can be discharged from intensified surveillance. Other women will “have an increased risk” and suddenly have access to intensified surveillance. However, the goal is to improve the efficiency of the intensified surveillance program. Therefore, it is very important to evaluate the outcome regularly and also to evaluate the results economically (55).

There is a need for fundamental schemes for risk communication and also to develop practical counseling aids. This is where the RISIKOLOTSE.DE project comes in (Figure 13). We are planning to develop an online platform which will provide information and tools that will allow the breast cancer risk to be calculated, understood and evaluated (18).
In conclusion our research shows that an individualized risk adapted screening strategy is feasible and of advantage in clinical practice. Our research could contribute to the evidence required to overcome any barriers associated with replacing the current screening guidelines with a more focused individualized approach, which requires medical counseling. It forms the basis for guidance required by political and funding institutions to assess the cost/benefit of implementing Mammography Screening 2.0 as an alternative to the current “all inclusive” age related screening guidelines.
Abstract

There has been a lot of controversy about the current mammography screening program. Screening recommendations for breast cancer in Germany are currently based solely on age and gender of an individual, despite the fact that additional genetic and non-genetic factors are known to influence cancer risk. Recent breast cancer risk models include these factors. They would allow the implication of a risk adapted screening approach, “Mammography Screening 2.0”. On the one hand there is a need to intensify diagnostic procedures for women at higher risk; on the other hand it is desired to avoid unnecessary diagnostic procedures in women who are unlikely to develop breast cancer. The aim is to improve the efficiency of the screening program and to help guide screening decisions by patients’ individual risk profiles and preferences.

However, the implementation of such a strategy faces new challenges, such as the choice of the adequate prediction model, the interpretation of the results, and the ways to communicate the risks.

We could show that currently used risk models, such as IBIS and BOADICEA, are well calibrated and conclusively already provide a useful perspective for individualized screening as they perform more effectively than considering age alone. Our ultimate goal is to categorize women according to their risk of breast cancer as accurately as possible, based on their profile of genetic and non-genetic risk factors, and to recommend a more individualized screening program. However, at the moment we do not have an international agreement how best to define for breast cancer the “high risk” group. We recommend that physicians use time periods of briefer duration, such as 5 or 10-year risks to identify women at high risk.

Appropriate risk communication in Mammography Screening 2.0 will present a new but interesting challenge for physicians. There is a need to implement shared decision-making in routine medical practice, and also a need for fundamental schemes for risk communication.

In conclusion our research shows that an individualized risk adapted screening strategy is feasible and of advantage in clinical practice. Our research could contribute to the evidence required to overcome any barriers associated with replacing the current “all inclusive” age related screening guidelines with a more focused individualized approach, which requires medical counseling.
List of Publications

Research articles as First or Last Author


**Research articles as Co-Author**


Published articles


References


73. Welch HG, Passow HJ. Quantifying the benefits and harms of screening mammography. JAMA internal medicine. 2014;174(3):448-54.

