Likelihood of early detection of breast cancer in relation to false-positive risk in life-time mammographic screening: population-based cohort study

J. D. M. Otten1*, J. Fracheboud2, G. J. den Heeten3, S. J. Otto2, R. Holland3, H. J. de Koning2, M. J. M. Broeders1,3 & A. L. M. Verbeek1

1Department for Health Evidence, Radboud University Medical Centre, Nijmegen; 2Department of Public Health, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam; 3National Expert and Training Center for Breast Cancer Screening, Nijmegen, the Netherlands

Received 5 December 2012; revised 28 March 2013; accepted 10 May 2013

Background: Women require balanced, high-quality information when making an informed decision on screening benefits and harms before attending biennial mammographic screening.

Patients and methods: The cumulative risk of a false-positive recall and/or (small) screen-detected or interval cancer over 13 consecutive screening examinations for women aged 50 from the start of screening were estimated using data from the Nijmegen programme, the Netherlands.

Results: Women who underwent 13 successive screens in the period 1975–1976 had a 5.3% cumulative chance of a screen-detected cancer, with a 4.2% risk of at least one false-positive recall. The risk of being diagnosed with interval cancer was 3.7%. Two decades later, these estimates were 6.9%, 7.3% and 2.9%, respectively. The chance of detection of a small, favourable invasive breast cancer, anticipating a normal life-expectancy, rose from 2.3% to 3.7%. Extrapolation to digital screening mammography indicates that the proportion of false-positive results will rise to 16%.

Conclusion: Dutch women about to participate in the screening programme can be reassured that the chance of false-positive recall in the Netherlands is relatively low. A new screening policy and improved mammography have increased the detection of an early screening carcinoma and lowering the risk of interval carcinoma.

Key words: breast neoplasm, cumulative risk, false-negative reactions, false-positive reactions, mammography, mass screening

Introduction

Early detection is still considered a cornerstone in the fight against mortality in many carcinomas and has been extensively evaluated in organised service screening mammography. From a population-based perspective, participation in a breast cancer screening programme is beneficial for a few, screened women in whom breast cancer is detected in its preclinical phase; however, there are potential drawbacks for many. Gøtzsche et al. noted that in the programmes they investigated, the drawbacks were neither fully nor correctly communicated to newly invited women. [1] Women need balanced, high-quality information to make an informed decision on the benefits and harms of participation. [2] This necessitates analysis of screening programmes on both the benefits and harms of long-term participation in relation to recall rates. Life-time risks of different screening outcomes can be calculated using empirical data from the long running service screening programme in Nijmegen, the Netherlands.

Methods

Setting

The Nijmegen population-based (invitational) screening programme started in 1975. More than 20 000 women aged 35–65 years were offered a screening examination once every 2 years. In 1989, the nationwide screening programme was introduced, and the target age-group of the Nijmegen programme was gradually limited to comply with the national policy of actively inviting women aged 50–69. The upper age limit was extended to 75 years in 1997. We used individual data from women now aged ≥75, who joined the programme when aged 50, and thus have undergone 13 consecutive examinations.

Screening mammograms are developed on site and read by two radiologists later, independently. In the Netherlands, women with a positive screening examination are referred to their general practitioners (GPs) and subsequently to a hospital of their choice for further assessment. Further diagnostic work-up is not an integrated part of the screening programme, but is carried out in general or academic hospitals. We use the commonly
used international term ‘recall’ for a case where a woman is referred to the GP following screening.

Women participating in the screening programme were asked to give written informed consent for their data to be used for evaluative purposes. Since the start of the project, data have been collected on diagnostic follow-up of participants with a positive screening outcome from the screening organisation, from GPs and the two regional hospitals. Data collection on women with an interval carcinoma is based on information from the screening organisation (reports from re-invited women themselves and from GPs), the nationwide pathology network and registry of histopathology and cytopathology, and linkage with the Regional Cancer Registry from 1989 onwards.

**definition of outcome measures**

In this study, we classified the benefits and harms based on the following four outcome measures. Beneficial outcomes are the detection of (small) screen detected cancers. Harmful outcomes are a false-positive result and/or an interval cancer. Based on the available data, it is not possible to quantify the extent of overdiagnosis and other adverse effects such as the risk of death due to breast cancer from radiation exposure, pain and discomfort.

First, a positive screening test is defined as a suspicious lesion on the mammogram, usually on the basis of consensus among two screening radiologists. A suspicious lesion required recall for further diagnostic follow-up to identify the possible presence of a malignancy.

Second, a true-positive result (i.e. screen-detected breast cancer) was based on histologically and/or cytologically proven breast cancer confirmed during the woman’s clinical work-up in the 6 months following a positive screening examination. Third, a positive screening examination was considered false-positive if the assessment was negative and no breast cancer (invasive or non-invasive) was diagnosed during the first year after screening. A false-positive result is counted, even when the assessment was negative and during the interval period to the next screening an interval cancer was diagnosed for another lesion (see the definition of false-negative result). Invasive clinical work-up comprised fine needle aspiration cytology, core needle biopsy or surgical biopsy. Lastly, we defined a false-negative result (so-called interval cancer) when breast cancer was diagnosed after a negative screening examination or assessment, in the interval before the next screening, that is, within 2 years.

**data analysis**

The long-term risk of recall (overall, false-positive recall and invasive work-up for false-positive recall) and breast cancer detection (overall, invasive cancer and tumour size <15 mm), was estimated using Elveback’s formula [3], in which the number of screening examinations was used instead of calendar time (see supplementary appendix, available at *Annals of Oncology* online). For interval cancer, we corrected the estimate for withdrawals. This adjusted approach is a result of the fact that these examinations are not discrete events, because interval cancers are diagnosed within the time interval of 2 years between two screening examinations. We used the Greenwood formula [4] to calculate the standard error and the corresponding 95% confidence interval of the cumulative estimated risks across potential 13 screening examinations.

The long-term likelihood of screening outcomes was calculated using a sequence of probabilities for the women who participated in consecutive examinations. After a false-positive recall, the possibility of breast cancer diagnosis (interval or screen-detected) as outcome remains. Women were censored when they missed one examination, or at the time of breast cancer diagnosis.

To study changes in outcomes over time, we compared historical data with that from a more recent calendar period. We defined two cohorts of screening participants.

The historical cohort contained invited women aged 48–52 at their first screening examination in 1975–1976 (N = 3,539) who could therefore have potentially participated in 13 screening rounds until the age of 75. Because the invitation policy was adjusted in 1989, women who reached the age of 70 (11th examination) before 1998 were no longer actively invited for screening. Women aged ≥70 were re-invited only if they had attended the previous screening examination. Therefore, the numbers of attendees decreased rapidly with age after the 11th invitation. To report on the long-term risk for 13 examinations, we estimated the risk for the 11–13th examination using the last observation carried forward approach, assuming that the risk for the 70–75 age group would not change much in comparison to the group aged 68–69.

Because of the change in recall policy and improved screening quality over time, figures from the historical cohort could not be used to estimate current long-term risks. We therefore assembled a second (current) cohort of women invited from 1997 onwards (N = 7,669) aged 49–51 at their initial screening examination, who currently have undergone 5 consecutive examinations. We used the Cox proportional hazards model to extrapolate the observed data from the five examinations to long-term risks.

**results**

More than 11 000 women had their first screening examination around the age of 50. In the historical cohort, 3,539 women were screened in 1975–1976, and of these, 2,883 accepted the invitation for a second examination (round) (see supplementary Table S1, available at *Annals of Oncology* online). The number participating in successive examinations declined to 937 in the 16th round. Thereafter, participation dropped to 80 in the 13th round, because of the lack of an active invitation policy for women over 70 years (see the Methods section).

During the 24-year screening period (13 biennial screening examinations), 157 from a total of 2,883 participants were referred once and 4 participants twice, resulting in 74 women with screen-detected breast cancer. Forty-eight women were diagnosed with an interval cancer.

Table 1 shows that the chance of being referred at least once after 13 rounds was 9.3% (95% CI 7.8% to 10.7%). The cumulative chance of screen-detected breast cancer and a false-positive recall were 5.3% (95% CI: 4.1% to 6.5%) and 4.2% (95% CI: 3.3% to 5.1%), respectively.

Figure 1 depicts the long-term chance of being detected with breast cancer or the risk of being referred false-positively at least once for the historical cohort. At the initial examination, the risk of a false-positive recall was much higher than the risk of a true-positive recall (1.13% versus 0.59%). In the long-term, the risk of a true-positive recall was higher (4.2% versus 5.3%, respectively). The risk of interval breast cancer is stable across the screening examinations and accumulates in the long run to 3.7%; 95% CI: 1.5% to 5.8% (Figure 1). The ratio of screen-detected versus interval breast cancer is ~3:2.

In the current cohort, more women were recalled than in the historical cohort. The average recall rate for a subsequent screening examination doubled from 6.3 (historical) to 11.1 per 1000 screenees in the current cohort, and the positive predictive value of the screening mammography decreased from 50% to 35%. The change in recall policy resulted in a higher detection rate, increasing at subsequent screening from 3.2 per 1000 for the historical cohort to 3.8 per 1000 in the current cohort. Interval cancer rates decreased from 2.3 per 1000 to 1.9 per 1000.
As a result, the cumulative chance (Table 1) of being recalled increased from 9.3% to 14.5% (95% CI: 11.5% to 17.5%), leading to an increase in the risk of a false-positive recall from 4.2% to 7.3% (95% CI: 5.5% to 9.0%), but a higher cancer detection from 5.3% to 6.9% (95% CI: 4.2% to 9.6%).

The chance of having a small (<15 mm) invasive tumour detected increased from 2.3% (95% CI: 1.5% to 3.2%) to 3.7% (95% CI: 0.9% to 6.4%). Most of the invasive work-up for the false-positive recalls took place in the first round (0.8% to 0.9%), which is approximately one-third of the total risk (2.4% to 2.6%).

The risk of a woman being confronted with interval cancer decreased from 3.7% (95% CI 1.5% to 5.8%) to 2.9% (95% CI 1.5% to 4.2%) in the current cohort. The ratio of screen-detected versus interval breast cancer changed from 3:2 to 5:2.

discussion

Based on the long-term data from the historical cohort, after 13 consecutive screening examinations the cumulative risk for a woman of having at least one false-positive mammogram was 4.2%; the chance of being detected with breast cancer was 5.3%. We found the highest risk for a false-positive recall in the initial screening and a gradual increase of the cumulative risk thereafter (more strongly in the first four rounds). In contrast, the increase in the cumulative likelihood for detection of cancer at screening is more pronounced as the number of examinations rises. The high risk for recall seen at initial screening is reasonable because of the prevalent cases detected during the initial examination and the absence of a previous comparison mammogram (more false-positives). Furthermore, women at younger ages have proportionally denser breast tissue, which diminishes gradually with increasing age, and causes more false-positive recalls at the beginning of the series than at the end. The incidence of breast cancer increases with age, which is corroborated by the increasing risk for screen-detected breast cancer.

The occurrence of interval cancer is stable over the 13 examinations, and the cumulative risk was lower for the current cohort (2.9% versus 3.7%). The ratio in the long-term risk of

### Table 1. Cumulative chance for recall and breast cancer diagnosis for a 50 year old woman based on 13 screening examinations

<table>
<thead>
<tr>
<th>Long-term chances early period* (95% CI)</th>
<th>Extrapolation of the chances in current cohort† (95% CI)</th>
<th>Expected figures in digital mammography§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>9.3% (7.8–10.7%)</td>
<td>14.5% (11.5–17.5%)</td>
</tr>
<tr>
<td>Screen-detected breast cancer</td>
<td>5.3% (4.1–6.5%)</td>
<td>6.9% (4.2–9.6%)</td>
</tr>
<tr>
<td>Invasive</td>
<td>4.4% (3.2–5.7%)</td>
<td>5.5% (3.1–7.9%)</td>
</tr>
<tr>
<td>Invasive cancer &lt;15 mm</td>
<td>2.3% (1.5–3.2%)</td>
<td>3.7% (0.9–6.4%)</td>
</tr>
<tr>
<td>Interval cancer</td>
<td>3.7% (1.5–5.8%)</td>
<td>2.9% (1.5–4.2%)</td>
</tr>
<tr>
<td>False-positive recall</td>
<td>4.2% (3.3–5.1%)</td>
<td>7.3% (5.5–9.0%)</td>
</tr>
<tr>
<td>Invasive work-up 1st exam</td>
<td>0.8% (0.5–1.1%)</td>
<td>0.9% (0.7–1.1%)</td>
</tr>
<tr>
<td>Invasive work-up 13 examinations</td>
<td>2.4% (1.7–3.1%)</td>
<td>2.6% (1.7–3.4%)</td>
</tr>
</tbody>
</table>

95% CI = 95% Confidence Interval.
*Early period (historical cohort): first screening in 1975, extrapolation for 13 examinations are based on 10 examinations.
†Current cohort: first screening examination in the period 1997–2006, extrapolation for 13 examinations are based on observations of 5 examinations from the early and current screening period.
‡Invasive work-up: invasive clinical examination like fine needle aspiration cytology, core needle biopsy or surgical biopsy.
screen-detected breast cancer versus interval cancer changed positively for the current cohort.

Several publications (see supplementary Table S2, available at Annals of Oncology online) have addressed the nature and magnitude of the cumulative false-positive risk in screening mammography and are already extensively discussed [13, 14]. The variations between the organized screening programmes from supplementary Table S2, available at Annals of Oncology online are presented in Figure 2. The different scenarios shown are based on false-positive rates originating from the IARC Handbooks of Cancer Prevention [15]. We also added the false-positive risk over time found for the UK triennial breast cancer screening programme, based on 6.97% false-positive recall at the first screening and 3.52% at subsequent screenings [16]. Although different protocols or women’s characteristics directed the cumulative risk of a false-positive examination [7, 11, 13], the main influencing factor is the tendency to interpret mammograms as abnormal, which is high in the United States and low in both the Dutch nationwide and Fynn’s regional screening programme. [10, 17] As false-positive recall is a direct consequence of the balance between the recall rate and the detection of breast cancer, the most important consequence of high recall rates is the disproportionate number of false-positive rates, while cancer detection rate levels off. [18] This is seen in supplementary Table S2, available at Annals of Oncology online, in which the detection rate varies by a factor of 2 at the most, whereas the variation in the false-positive rates is almost a factor of 10.

![Figure 2. Cumulative risk of false-positive recall for various scenarios (rates originating from IARC Handbooks of Cancer Prevention [20], studies, cf. supplementary Table 2, available at Annals of Oncology online)](image-url)
Due to an altered recall policy, the cumulative risk of a false-positive recall increased to over 7% in the current cohort of the Nijmegen programme. The introduction of digital mammographic screening will lead to even higher changes in recall rates. Based on the initial and subsequent digital false-positive rates from the Dutch breast cancer screening programme [5], we estimated the cumulative risk using Elveback’s formula. The cumulative false-positive rate will now increase from 7% in analogue screening to 16% after 13 examinations with full-field digital mammography (see NL-u, in Figure 2). This 16% false-positive risk is still lower than that found in analogue screening, for instance, the weighted pooled estimate of 19.7% from three studies based on 10 biennial analogue screening rounds calculated by Hofvind et al. [14]. Compared with other digital screening programmes, the estimated cumulative risk of a false-positive recall in the Netherlands (16%) remains much lower than in other European countries [19, 20], in which, the cumulative risk (estimated in the same manner; see above) ranges from 30% to 40%.

In a recent article, Gøtzsche et al. stated that women in the UK are still poorly informed about the side-effects of breast cancer screening. [1] In our study, we have quantified specific effects which are relevant to the provision of clear information to potential screenes, for example the long-term risk of false-positive recall and diagnosis of interval cancer, and we have related these potential harmful effects to the potential beneficial effects of early detection of cancers.

Another issue noted by Gøtzsche was the extent to which screening saves lives. In our study, the long-term risk of screen-detected invasive breast cancer <15 mm is estimated to be 3.7%. As previously demonstrated by Otten et al. [21], patients with small invasive cancer (tumour size <15 mm) have the same life expectancy as women in the target population invited for screening. In combination with adequate therapy, 67% (=3.7%/5.5%, Table 1, current cohort) of all women with a screen-detected invasive breast cancer will have a normal life expectancy.

Our study setup does not enable measurement of overdiagnosis, one of the drawbacks of screening, but based on van Gelder’s study [22] the extent of overdiagnosis in the Dutch screening programmes seems to be limited to 9.7%. One other possible limitation of our study is that, due to the number of examinations over time, both age and time related changes may have influenced the risks. When estimating the risk for 13 examinations, we assumed that the time-related changes in the earlier period could also be expected in the near future. Should the (technical) improvement of the screening test have levelled off in recent years, the extrapolated risks from the current period to 13 examinations will be overestimated. On the other hand, should screening tests further improve, for example due to the use of or advances in digital mammography, the cumulative risks will have to be re-estimated for different screening outcomes.

Though the low recall rate in the Nijmegen programme, the detection rate for subsequent examinations are comparable with, e.g. those found by the Barcelona programme and somewhat lower than in the Scandinavian programmes. Moreover, despite the low recall policy, the Nijmegen screening programme has shown considerable breast cancer mortality reduction, which varies from 28% to 65%, depending on the calendar period. [23]

Conclusion
We have been able to provide data on a programme with low recall rates and place them in a broader perspective. At this moment, the long-term risks of unintended effects for dedicated participants in this specific long running screening programme appear to be within the reasonable limits. Although the recall rate has increased substantially in more recent years, it can still be considered low in an international perspective.

We conclude that Dutch women about to participate in the screening programme can be reassured that the chance of a false-positive recall in the Netherlands is relatively low, and that screen detection of early breast cancer has improved at the cost of a lower risk of interval cancers.

Disclosure
The authors have declared no conflicts of interest.

References
Obesity, diabetes, and survival outcomes in a large cohort of early-stage breast cancer patients

S. Jiralerspong1,2*, E. S. Kim2,3, W. Dong4, L. Feng4, G. N. Hortobagyi2 & S. H. Giordano2

1Lester and Sue Smith Breast Center, Department of Medicine, Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston; 2Department of Breast Medical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston; 3James P. Wilmot Cancer Center, Department of Medicine, University of Rochester, Rochester; 4Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, USA

Background: To determine the relationship between obesity, diabetes, and survival in a large cohort of breast cancer patients receiving modern chemotherapy and endocrine therapy.

Patients and methods: We identified 6342 patients with stage I–III breast cancer treated between 1996 and 2005. Patients were evaluated according to body mass index (BMI) category and diabetes status.

Results: In a multivariate model adjusted for body mass index, diabetes, medical comorbidities, patient- and tumor-related variables, and adjuvant therapies, relative to the normal weight, hazard ratios (HRs) for recurrence-free survival (RFS), overall survival (OS), and breast cancer-specific survival (BCSS) for the overweight were 1.18 [95% confidence interval (CI) 1.02–1.36], 1.20 (95% CI 1.00–1.42), and 1.21 (95% CI 0.98–1.48), respectively. HRs for RFS, OS, and BCSS for the obese were 1.13 (95% CI 0.98–1.31), 1.24 (95% CI 1.04–1.48), and 1.23 (95% CI 1.00–1.52), respectively. Subset analyses showed these differences were significant for the ER-positive, but not ER-negative or HER2-positive, groups. Relative to nondiabetics, HRs for diabetics for RFS, OS, and BCSS were 1.21 (95% CI 0.98–1.10), 1.39 (95% CI 0.98–1.49), 1.39 (95% CI 0.98–1.49), respectively.

Conclusions: In patients receiving modern adjuvant therapies, obesity has a negative impact on RFS, OS, and BCSS; and diabetes has a negative impact on RFS and OS. Control of both may be important to improving survival in obese and diabetic breast cancer patients.

Key words: body mass index, breast cancer, diabetes, obesity, prognosis, survival outcomes

Introduction

Obesity has reached epidemic proportions, with about two-thirds of the US population and one-third of the world population being either overweight or obese [1]. Obesity is an established risk factor for developing postmenopausal breast cancer [2–4]. Recent data suggest that obesity is also associated with worse survival outcomes (prognosis) in both pre- and postmenopausal breast cancer [5–8]. Diabetes has also reached epidemic proportions, with 17%–20% of the US population over age 65 being diabetic [9]. Diabetes is correlated with an increased risk of breast cancer [10], as well as with poorer breast cancer outcomes [11–13].

Obesity and diabetes share overlapping etiologies. Patients with both conditions have elevated serum insulin levels and show evidence of a chronic inflammatory state [14]. Both conditions are associated with worse breast cancer outcomes in...