General introduction

Breast cancer is the most frequently diagnosed cancer in women worldwide. Also in the Netherlands, breast cancer is an important threat for public health as the incidence is among the highest in Europe, and still rising. As in many other countries, the Netherlands has introduced a population based screening program, with the aim to reduce breast cancer mortality by detecting the disease at an early stage.

Screening can only be effective if certain criteria are met, such as a sufficiently high attendance, referral- and detection rate, a sufficiently high sensitivity and specificity and a relatively low rate of interval cancers and false positive screening results. Screening programs are evaluated based on these parameters, and these parameters have changed over time. The most important changed since the introduction of screening mammography was the increase in recall rate, and thereby an increase in cancer detection rate, screening sensitivity and false positive rate.

Screening outcome can be influenced by a number of parameters including patient (e.g. breast density, previous breast surgery, breast cancer prevalence), imaging (i.e. screen film or digital mammography) and screening situation related parameters (i.e. reading strategy, radiologist performance).

This thesis will explore several factors of the breast cancer screening program in a Southern screening region of the Netherlands with the aim to improve the effectiveness of screening. The main focus of this thesis will be the evaluation of patients with a repeated false positive recall at the transition to digital screening, radiologist’s performance at non-blinded double reading and to determine the best reading strategy at double reading.

Re-attendance at biennial screening mammography following a repeated false positive recall

The first objective of this thesis was to determine the re-attendance rate at screening mammography after a single or a repeated false positive recall (chapter 2). We found that the re-attendance rate for women who had experienced a single false positive recall is significantly worse compared to patients with a negative screen at prior screening mammography (65.4% versus 93.2%). Re-attendance after a repeated false positive recall is even worse (52.1%), especially if both recalls comprised the same mammographic lesion (44.3%).

The second purpose of chapter 2 was to determine the effect of the transition from screen film to digital mammography on the proportion and screening outcome of women recalled twice for the same lesion. Moreover, we assessed the possible influence of comparison with scanned in priors instead of hard copy priors at digital screening on the proportion of women recalled twice for the same mammographic abnormality. We found that, during the first screening round at digital screening, a significantly larger proportion of recalls included women who had been recalled twice for the same lesion and breast cancer was significantly less often diagnosed in these
women than at screen film mammography, with a concomitant lower positive predictive value (PPV) of recall. In this population, subsequent digital screens were compared with previous digitized analogue screens. Blinded review showed that the availability of the screen film hard copies would have reduced the number of repeatedly recalled women at digital mammography by almost 40%. None of these women were ultimately diagnosed with breast cancer.

We therefore suggest that, in case a woman attends digital screening for the first time after previous screen film mammography (e.g., screening programs currently converting to digital mammography and women who have skipped the digital screening rounds following their latest SFM screen), the screening radiologist should have the opportunity to compare with the previous hard copy screen film examination if repeated recall is considered.

Variations in screening outcome among pairs of screening radiologists at non-blinded double reading

Substantial inter-observer variability in screening mammography interpretation among screening radiologists has well been documented. However, screening results of pairs of screening radiologists have not yet been published. In chapter 3 we determined the variations in screening performance among unique pairs of screening radiologists at non-blinded double reading. We expected that the addition of a second reader would decrease these variations in screening outcome, especially among a group of screening radiologists who attend quality assurance sessions at the same regular intervals. Nevertheless, our results proved otherwise. Variations in screening performance significantly varied among 26 pairs of screening radiologists with respect to recall rate, program and mammographic sensitivity and PPV of recall. This stresses the importance of monitoring screening results on a local scale.

Blinded versus non-blinded double reading

European guidelines consider double reading by screening radiologists as standard of reference for the assessment of screening mammograms as this reading strategy significantly increases the cancer detection rate when compared to single reading. Double reading can be performed in a blinded (2nd reader is not informed about the 1st reader’s decision) or non-blinded fashion (2nd reader is informed about the 1st reader’s decision). In chapter 4.1 we determined the screening mammography outcome at blinded and non-blinded double reading in a prospective population based study. In this study women with discrepant readings between the two radiologists were always recalled for further analysis.

We found that blinded double reading almost doubled the proportion of discrepant readings among recalled women in our study, which resulted in a significantly higher recall rate and false positive rate. However, we found that blinded double reading yielded a significantly better program sensitivity than non-blinded double reading, as
well as a 14% higher (7.4 versus 6.5), although non-significant, cancer detection rate. We therefore advocate the use of blinded double reading instead of non-blinded double reading.

Diagnostic work-up after blinded and non-blinded double reading

In the previous chapter we advised the use of blinded double reading over non-blinded double reading. However, the disadvantage of blinded double reading is that it resulted in a higher false positive rate. It is known that the degree of distress after a false positive recall is related to the invasiveness of the assessment. Therefore we compared the diagnostic work-up after blinded and non-blinded double reading in chapter 4.2.

We found that blinded double reading resulted in significantly higher overall (17.4 versus 14.3) and benign biopsy rate (10.1 versus 7.7) than non-blinded double reading. This difference was caused by higher core needle biopsy (CNB) and stereotactic core needle biopsy (SCNB) rates. After non-blinded double reading, a larger proportion of women at benign work-up were evaluated merely by additional imaging. This negative side effect of blinded double reading should be taken into account when adopting this reading strategy.

Arbitration of all discrepant readings at blinded and non-blinded double reading

Blinded double readings results in discrepant readings. In the previous chapters all discrepant readings were recalled for further evaluation. The objective of chapter 4.3 was to determine the value of adding a third reader for arbitration of discrepant screening mammography assessments.

Discrepant screening mammography occurred more often at blinded (57.2%) than at non-blinded double reading (29.1%). Arbitration of all discrepant readings by a third reader would have improved the recall rate and positive predictive value (PPV) of recall at both reading strategies. However, it would have significantly decreased the program sensitivity at blinded double reading. There was no significant effect on the cancer detection rate at both reading strategies and the program sensitivity at non-blinded double reading. This study therefore showed that arbitration of all discrepant screening mammography assessments is a good tool to improve recall rate and PPV, but is not desirable as it reduces the program sensitivity at blinded double reading.

Characteristics of BI-RADS category 0 readings and the effect of arbitration of discrepant BI-RADS 0 readings on screening outcome

BI-RADS category 0 recalls represent lesions with a low suspicion for malignancy, requiring additional work-up. In chapter 4.4 we evaluated the characteristics of BI-RADS 0 lesions at blinded and non-blinded double reading and determined the
potential effect of arbitration of discrepant BI-RADS 0 recalls by a third reader on screening outcome.

We found that both at blinded and non-blinded double reading, 32.0% and 32.5% of recalls were assigned BI-RADS 0 with a positive predictive value of 7.2% and 6.8%, respectively. Compared to non-blinded double reading, BI-RADS 0 recalls at blinded double reading showed a higher discrepancy rate and false positive recall rate. Arbitration of discrepant BI-RADS 0 recalls would have significantly lowered the recall rate, without a decrease in cancer detection rate and program sensitivity. Arbitration would have significantly increased the PPV at blinded double reading. On that account, we advise arbitration of discrepant BI-RADS 0 recalls, at both blinded and non-blinded double reading.

General discussion

To maintain the effectiveness of breast cancer screening in the rapidly changing therapeutic landscape, it is relevant to continuously evaluate the screening program and try to maximize its benefits and minimize the drawbacks of mass screening. Based on the studies presented in this thesis, the following strategies may improve the effectiveness of the population based breast cancer screening program:

1. If a woman attends digital screening for the first time after previous screen film screening (i.e. screening programs currently converting to digital mammography and women who have skipped the digital screening rounds following their latest SFM screen) the screening radiologist should have the opportunity to compare with the previous hard copy SFM examination if repeated recall is considered.

2. Create specific pairs of radiologists in which radiologists with good to superior screening results are combined with radiologists with less optimal screening results or in which radiologists specialized in breast imaging are combined with radiologists who are not.

3. To detect suboptimal individual screening results or suboptimal results of pairs of screening radiologists, it is important to monitor screening results at a local and individual level.

4. Although a cost effectiveness analysis is still required, blinded double reading with arbitration of discrepant BI-RADS 0 recalls seems the optimal reading strategy to improve the effectiveness of the breast cancer screening program in the Netherlands.

Additionally, a personalized risk-based approach, especially for patients with dense breast tissue, may further improve the efficacy of breast cancer screening in the near future.