Strategies to improve breast cancer detection by digital screening mammography

Summary

Radiologist blinded double reading

The first objective of this thesis was to determine the impact of the second reader on screening outcome (**chapter 2**). We found that adding a second reader resulted in a significant increase in cancer detection rate (from 6.2 per 1000 screens to 7.0 per 1000 screens, p<0.001) at the expense of an acceptable increase in recall rate (from 3.0% to 3.6%, p<0.001) and false positive recall rate (from 23.4 per 1000 screens to 28.7 per 1000 screens, p<0.001). Based on these results we therefore favour blinded double reading over single reading.

In addition to the overall screening outcome, we also compared mammographic and tumour characteristics between first and second reader recalls. We found a larger proportion of BI-RADS 0 lesions in de group of women additionally recalled by the second reader and recalled lesions comprised larger proportions of asymmetries and architectural distortions. The majority of 82 cancers additionally detected by the second reader were small (86.9% T1a-c) invasive cancers, mostly of low histological grade (59.0% grade I).

Because of our study design, additional recalls by the second reader are by definition discordant recalls. However, the second reader might not agree with all recall decisions made by the first reader (i.e. there are also discordant readings among first reader recalls). In **chapter 3** we further analysed this and determined which mammographic and tumour characteristics led to concordant versus discordant readings. A discordant reading was defined as a difference in classification by two radiologists where one classified it as BI-RADS 1 or 2 (negative, i.e. no recall) and the other as BI-RADS 0, 4 or 5 (positive, i.e., recall). In case of a discordant reading, the woman in question was always recalled for further analysis. The majority of recalls (71.8%) were based on a concordant reading. The PPV of a concordant recall was higher compared with a discordant recall (23.5% vs. 10.0%, p<0.001). As we expected based on the results presented in **chapter 2**, the proportion of BI-RADS 0 recalls was significantly higher in the discordant reading group (75.7% vs. 56.3%, p<0.001) and discordant readings were more often an asymmetry or architectural distortion (21.8% vs. 13.2% and 9.3% vs. 6.8%, respectively; p<0.001). Invasive cancers detected through discordant recalls showed a more favourable (grade I) tumour grade (54.7% vs. 39.9%, p=0.022). There was also a trend towards smaller tumours with fewer cancers showing lymph node metastases (16.0% vs. 23.5%, p=0.117). Other tumour characteristics were comparable for both groups.

Value of technologist's reading and quality assurance sessions

Since the initiation of our screening programme, screening technologists have been encouraged to look for mammographic abnormalities. Their training not only focuses on mammography technique and positioning but also on breast anatomy and pathology (benign and malignant). After obtaining a mammogram, they review it and annotate any abnormality for which they consider recall necessary.

In **chapter 4.1** we determined whether this technologist reading could be used to decide if discordant readings at radiologist blinded double reading should be recalled. We found that the proportion of breast cancer cases among discordant recalls was comparable for women with or without a positive technologist reading, irrespective of BI-RADS classification (5.0% vs. 4.6% for BI-RADS 0 recalls, p=0.8; and 28.9% vs. 25.4% for BI-RADS 4 or 5 recalls, p=0.5). This indicates that, in deciding whether a discordant reading should be recalled, assessment by a technologist does not provide a significant discriminating ability. When also considering the, borderline significant, decrease in cancer detection rate, there appears to be no added value of using the technologist's assessment of screening mammograms in arbitrating discordant readings.

For the purpose of quality assurance and continued training, technologists attend 6-weekly sessions where a coordinating screening radiologist discusses all mammograms that were considered suspicious by the technologists but were not recalled by the screening radiologists. Additional tumours are detected through these quality assurance sessions and in **chapter 4.2** we determined the frequency and characteristics of these cancers. Over 466,000 mammograms were included, resulting in 14,142 recalls and 3156 screen-detected cancers. During our 8-year inclusion period, technologists would have recalled 11,627 women with a negative radiologist screen (i.e., BI-RADS 1 or 2). All these mammograms were discussed at quality assurance sessions, resulting in 85 recalls and 26 screen-detected cancers (0.8% of all screen-detected cancers). Most of these cancers were invasive and larger than 10 mm in size, therefore probably not reflecting overdiagnosis. Sensitivity of quality assurance sessions for additional cancer detection was 52% (26 of 50; 95% confidence interval: 38%, 66%). We found no significant differences in mammographic and tumour characteristics compared with the 3130 screen-detected cancers detected through standard radiologist double reading. We also compared the cancers detected through quality assurance sessions with interval cancers and screen-detected cancers with a technologist positive screen for the same abnormality in a previous screening round (i.e., 'missed' at quality assurance sessions) and also found no significant differences. The role of these sessions in additional cancer detection appears to be

limited. However, they serve many other purposes, such as continued training, discussing interesting and instructive cases and receiving feedback on performance.

Frequency and characteristics of additional breast abnormalities following screening mammography

In clinical work-up after recall, additional breast cancers may be detected in the contralateral and/or ipsilateral breast. We studied the frequency and outcome of additionally detected contralateral (**chapter 5.1**) and ipsilateral (**chapter 5.2**) breast abnormalities following recall. We found that 3.2% of recalled women underwent assessment of a contralateral, non-recalled breast abnormality and 6.4% of recalled women was assessed for a ipsilateral, non-recalled lesion. In both groups, the additional lesion was in most cases first detected by review of the screening mammogram or additional mammographic views, with or without breast tomosynthesis. Since tomosynthesis is increasingly being used in clinical and, to a lesser extent, screening setting, it is expected that the number of additional lesions detected after recall will increase. Women with an additionally detected breast abnormality more often had high breast density (ACR III + IV), 38.8% vs. 23.8% for contralateral lesions ($p<0.001$) and 29.0% vs. 23.8% for ipsilateral lesions ($p=0.06$). The majority of additionally detected lesions were benign and mostly comprised cysts and fibroadenomas. Additionally detected lesions proved to be malignant in 19.0% (ipsilateral) and 15.5% (contralateral) of cases, respectively. Most additional ipsilateral malignancies were part of a multifocal or multicentric screen-detected tumour. Positive predictive value of biopsy was higher for recalled lesions compared with non-recalled lesions. However, in case of an additional ipsilateral lesion, PPV of biopsy was higher in women with a malignant recalled lesion compared to benign recalled lesions. The latter implicates that a low biopsy threshold is warranted in case of a malignant index lesion.