BREAST CANCER MISSED AT SCREENING; HINDSIGHT OR MISTAKES?

Abstract

Purpose: To investigate radiologists' interpretation scores of screening mammograms prior to diagnosis of screen-detected and interval breast cancers retrospectively classified as missed or true negative.

Methods: We included data on radiologists' interpretation scores at screening prior to diagnosis for 1223 screen-detected and 1007 interval cancer cases classified as missed or true negative in an informed consensus-based review. All prior screening examinations were independently scored 1-5 by two radiologists; score 1 by both was considered concordant negative, score ≥2 by one radiologist discordant, and score ≥2 by both concordant positive. We analyzed associations between interpretation, review categories, mammographic features and histopathological findings using descriptive statistics and logistic regression.

Results: Among screen-detected cancers, 31% of missed and 10% of true negative cancers had discordant or concordant positive interpretation at prior screening. The corresponding percentages for interval cancer were 21% and 8%.

Age-adjusted odds ratio (OR) and 95% confidence interval (CI) for missed screen-detected cancer was 3.8 (95% CI: 2.6-5.4) after discordant and 5.5 (95% CI: 3.2-9.5) after concordant positive interpretation, using concordant negative as reference. Corresponding ORs for missed interval cancer were 3.0 (95% CI: 2.0-4.5) for discordant and 6.3 (95% CI: 2.3-17.5) for concordant positive interpretation.

Asymmetry was the dominating mammographic feature at prior screening for all, except concordant positive screen-detected cancers where a mass dominated. Histopathological characteristics did not vary statistically with interpretation.

Conclusions: Most cancers were interpreted negatively at screening prior to diagnosis. Increased risk for missed screen-detected or interval cancer was observed after positive interpretation at prior screening.

Highlights

- Screen-detected and interval cancers were retrospectively classified missed or true
- Missed cancers were more often interpreted positively at prior screening than true
- Ten percent of missed screen-detected cancers were recalled at prior screening
- Asymmetry was dominating mammographic feature at prior screening in missed cases
- Histopathological characteristics were not associated with interpretation scores

Abbreviations

OR – Odds ratio

CI – Confidence interval

BI-RADS - Breast Imaging – Reporting and Data System

SD – Standard Deviation

IQR – Interquartile range

AI – Artificial Intelligence

Key words

breast; neoplasms; mass screening; mammography; female

Introduction

Mammographic screening has been implemented widely to reduce mortality from breast cancer by early detection [1, 2]. High quality of images as well as interpretation performance by screen readers are prerequisites to succeed [1]. The interpretation procedure varies across screening programs, with screen-reading by one or two radiologists with or without consensus/arbitration to decide whether to recall for further assessment as applied strategies [3]. Studies have proven higher rates of screen-detected cancer in programs with double versus single reading, and European guidelines and the European Commission Initiative on Breast Cancer recommend double reading [4-7].

Breast cancer among screening attendees is mainly diagnosed after an abnormal screening examination, but approximately one fourth presents after a negative screening episode and before the next scheduled screening, as interval cancer [8]. The majority of these cancers were not visible at prior screening, generally classified as true negative. However, some cases are classified as missed or false negative, indicating visible findings at prior mammograms when retrospectively reviewed [9-11].

Studies have shown an increased risk of screen-detected and interval cancer among women with a positive interpretation at prior screening, both cases dismissed at consensus as well as those recalled for further assessment, which turned out negative [12, 13]. However, the initial interpretation scores of the mammograms prior to diagnosis for cancer cases retrospectively classified as missed and true negative are to our knowledge, not published. Knowledge about these aspects may be important in optimization of mammographic screening including improvement of early detection.

The population based mammographic screening program in Norway, BreastScreen Norway, holds a unique database of detailed information on radiologists' interpretation scores and screening outcome. In the present study, we combined this information with data from a nationwide radiological review of prior screening and diagnostic mammograms of screen-detected and interval cancer [10, 11].

The aim of the study was to investigate the radiologists' interpretation scores of prior screening mammograms for cancers retrospectively classified as missed or true negative. We also explored the associations of interpretation scores with mammographic features as well as histopathological characteristics.

Material and methods

The Data Protection Officer for the Cancer Registry of Norway and the Heads of Department and/or research administration at the local breast centers approved this retrospective study (PVO approval number 2016/4696). The Cancer Registry Regulations waived the requirement to obtain written informed consent [14]. We received de-identified data from the Cancer Registry for analyses.

BreastScreen Norway invites women aged 50-69 to two-view biennial mammographic screening, and is described in detail elsewhere [15]. All screening data, including results from histopathological reports, are registered and stored in a national screening database at the Cancer Registry, which administers the program.

Study sample and characteristics

The study sample included 1223 women with screen-detected and 1007 women with interval cancer, all diagnosed in BreastScreen Norway 2004-2016 (Figure 1). All women with screen-detected cancer had a prior screening examination two years earlier, and all with interval cancer had a prior screening examination within the past two years before diagnosis. A panel of five experienced breast radiologists reviewed the mammograms from diagnosis and from the screening examination prior to diagnosis for these 2230 cancer cases. The review was fully informed, with all information and images from screening, diagnosis and recall assessment available, including histopathological data. The review panel classified the cancers according to findings on prior screening mammograms; cancers with visible (clear or subtle, but specific) abnormalities on prior screening mammograms at the later cancer site were classified as missed, whereas cancers with no visible abnormalities on prior screening mammograms at the later cancer site were classified as true negative (Figure 1). The decisions were consensus-based, and in case of multifocality or bilateral disease, the largest tumor was classified. The review is described in more detail elsewhere [10, 11].

Interpretation scores

The screening procedure in BreastScreen Norway includes independent reading by two breast radiologists, who both assign each breast a score of 1 to 5 (1 = normal/benign; 2 = probably benign; 3 = intermediate suspicion; 4 = probably malignant; 5 = malignant). All examinations with a score ≥2 by one or both radiologists for at least one breast, are discussed in consensus

with two or more radiologists to decide whether to recall the woman. Recall may be omitted for all consensus cases regardless of score combinations. However, if an examination given a score ≥3 is to be omitted from recall, the radiologist who originally assigned the score should be informed about the decision, and have the opportunity to raise objections to the decision [15].

In this study, we included the interpretation scores for each breast given by the two radiologists at the screening examination prior to diagnosis. We defined a concordant negative interpretation as interpretation score 1 by both radiologists, discordant interpretation as score 1 by one radiologist and ≥2 by the other, and a concordant positive interpretation as score ≥2 by both radiologists. All analyses of interpretation scores were performed per breast (breast-based).

Mammographic features and histopathological characteristics

Data from the review included mammographic features at prior screening for missed cancers, classified according to the Breast Imaging – Reporting and Data System (BI-RADS) 5th edition (mass, calcifications, asymmetry, distortion or other findings) [16]. Histopathological data from the Cancer Registry's database for the screen-detected and interval cancers included histopathological type (ductal carcinoma in situ or invasive carcinoma), and for invasive carcinomas tumor diameter, histological grade (1-3), axillary lymph node involvement (positive/negative), and estrogen/progesterone receptor status (positive/negative). In case of multifocal or bilateral disease, only the largest was included in the analyses.

Statistical analyses

We presented age (years) at diagnosis as mean with standard deviation (SD). We performed descriptive analyses of the interpretation scores at prior screening by review classification categories for both the breast with later cancer and the contralateral breast. For the breasts with cancer, we also performed descriptive analyses of mammographic features at prior screening by interpretation scores. Lastly, we analysed histopathological tumor characteristics by interpretation scores. Recall at prior screening was calculated as the proportion of women recalled for further assessment due to mammographic findings, both in the breast with cancer separately, as well as both breasts (total recall). We presented categorical data as numbers and percentages and tumor diameter as median (mm) with the interquartile range (IQR). We tested for statistical significance by chi-square test, Fisher's exact or non-parametric tests as appropriate. Logistic regression was used to calculate the age-adjusted odds ratio (OR) with

95% confidence intervals (CI) for missed screen-detected and interval cancer for screening mammograms with discordant and concordant positive interpretations, using concordant negative as reference. All analyses were stratified by detection mode (screen-detected or interval cancer). We used SPSS statistics version 26 (IBM Corp.) for all analyses and a significance level of 0.05.

Results

Interpretation at prior screening for the breast later diagnosed with cancer

Among screen-detected cancers, 83% (1009/1223) had concordant negative interpretation at prior screening, 12% (146/1223) had discordant and 6% (68/1223) had concordant positive interpretation. Among interval cancers, 87% (875/1007) had concordant negative, 11% (111/1007) discordant and 2% (21/1007) concordant positive interpretation (Table 1).

Thirty-one percent (141/457) of screen-detected cancers classified as missed and 10% (73/766) classified as true negative, had a positive score (≥2) by one or both readers at prior screening. The corresponding percentages for interval cancer were 21% (84/396) for missed and 8% (48/396) for true negative (Table 1).

Age-adjusted OR with 95% CI for missed screen-detected cancer was 3.8 (2.6-5.4) after discordant interpretation and 5.5 (3.2-9.5) after concordant positive interpretation compared to concordant negative. The corresponding OR for missed interval cancer was 3.0 (2.0-4.5) for discordant and 6.3 (2.3-17.5) for concordant positive interpretation(Table 2).

The vast majority of mammograms with a positive interpretation at prior screening were given a score of 2; 77% among screen-detected cancers and 86% among interval cancers, no statistically significant differences between missed and true negatives (Table 3).

Interpretation of the contralateral breast

Ninety-five percent (1166/1223) of women with screen-detected and 94% (950/1007) of women with interval cancer had concordant negative interpretation for the breast without cancer at prior screening, no statistically differences between missed and true negative cancers (Table 1).

Review classification category and recall

Total recall at prior screening was 10.3% (126/1223) for screen-detected cancers and 4.2% (42/1007) for interval cancers. Recall due to findings in the breast later diagnosed with cancer was 8.2% for screen-detected and 3.0% for interval cancer; 14.9% for missed and 4.2% for true negative screen-detected cancer ($p<0.001$), and 4.0% for missed and 2.3% for true negative interval cancer ($p=0.18$) (Table 4).

Interpretation and mammographic features

The most frequent mammographic feature at prior screening for missed screen-detected cancers was asymmetry for those with concordant negative (43%, 136/315) and discordant interpretation (36%, 33/92). The most frequent feature among those with concordant positive interpretation was mass (49%, 24/49). For missed interval cancer, asymmetry was the most frequent mammographic feature at prior screening for all: 46% (143/312) for concordant negative, 40% (27/68) for discordant and 38% (6/16) for those with concordant positive interpretation (Figure 2).

Interpretation and histopathological characteristics

We observed no statistically significant differences in histopathological characteristics between concordant negative, discordant, or concordant positive screen-detected cancers. This also applied to interval cancers (Table 5).

Discussion

We found that most cancers, both missed and true negative, were interpreted negatively at screening prior to diagnosis. However, 31% of screen-detected cancers retrospectively classified as missed had a positive interpretation of their screening examination prior to diagnosis, indicating a potential for earlier diagnosis. Fifteen percent of the missed screen-detected cancers and 4% of the missed interval cancers were recalled at prior screening due to positive score in the same breast as later diagnosed with cancer. Asymmetry was the most frequent mammographic feature at prior screening for the missed cancers, except for concordant positive screen-detected cancers where mass was most frequent.

Interpretation and recall

The association between the radiologists' interpretation scores of the mammograms at screening prior to diagnosis and cancers classified as missed and true negative in a

retrospective informed review, is to our knowledge, not previously described. However, studies have demonstrated that women with a false positive screening result (recalled for further assessment that turned out to be negative) are at higher risk for later screen-detected or interval cancer [12, 17-19]. This is consistent with the recall rate of 10.3% at prior screening for screen-detected cancer in our study, which is substantially higher than the overall recall rate in BreastScreen Norway (3.2%) during the period 1996-2021 [15]. The higher proportion of positive interpretation scores in missed cancers and a recall rate of 16% for missed screen-detected cancers, may indicate that a noticeable number of these cancers were perceived, but interpreted as not sufficiently significant for recall, or recalled with a negative recall assessment. Information about location of the finding resulting in a positive interpretation was not available, thus, we were not able to state to what extent the interpretation corresponded to the location of the later cancer. However, in a previous retrospective study from BreastScreen Norway, 43% of interval cancers with a false positive screening result in the screening round prior to diagnosis were recalled for the same finding as later diagnosed as cancer [20]. Consequently, as women with a false positive screening result are shown to have higher risk for later screen-detected or interval cancer [18, 19], a personalized screening regime, e.g. more frequent screening, other screening tools or additional tests might be beneficial. However, such an approach must be balanced against the increased costs and need for resources.

The screening procedure in BreastScreen Norway is independent double reading with consensus to decide whether to recall, and examinations with discordant as well as concordant positive scores are discussed in consensus. However, recall may also be based on an either positive interpretation – which means that all women with a positive score by at least one reader are recalled for further assessment, without consensus/arbitration. In our study sample the latter strategy could potentially have reduced the number of missed cancer, as 31% of the screen-detected and 21% of the interval cancers classified as missed actually had a positive interpretation score for the breast later diagnosed with cancer by one or both radiologists at prior screening. The trade-off is, however, a marked increase in recall rate as well as false positive screening results.

Radiologists miss cancers at screen reading due to misperception, as illustrated by the relatively high proportion of concordant negative interpretation scores among the cancers retrospectively classified as missed. However, a fully informed consensus based radiological review yields the highest proportion of missed cancers [8, 21, 22]. The review situation is not

comparable to a real screening setting, in which radiologists interpret large batches of mostly negative mammograms with the accepted or expected recall rates and false positive rates of the program in mind. Organized testing and training, regular audits with feedback to the readers about their performance as well as test-sets with interactive feedback have a positive impact on reader sensitivity and the proportion of false positives [23-27].

Artificial intelligence (AI) has evolved as a promising tool in mammographic screening. In studies, performance of AI reading mammograms equals performance of average radiologists, and further improvement in performance is expected [28-30]. If AI may assist or even outperform radiologists in their ability to identify findings requiring a recall, an improved sensitivity and specificity of mammographic screening might be the outcome. In our study, most of the missed cancers were interpreted concordant negative at prior screening. Further, the vast majority of the prior mammograms with a positive interpretation at prior screening were given a score of 2 (probably benign). Hence, if AI is capable of detecting some of the negatively interpreted cancers, or at least raising the suspiciousness for cancer, the sensitivity might increase, followed by more cancers detected in an early stage.

Mammographic features and histopathological characteristics

An asymmetry is a one- or two-dimensional non-space occupying lesion, often representing normal breast tissue, whereas a mass is a convex, space-occupying biplane lesion [16, 31]. For cancers classified as missed, the proportion of masses at prior screening was higher for those with concordant positive scores than concordant negative scores, and vice versa for asymmetries. Hence, our results indicate asymmetry to be the most common feature missed at screening. This corresponds to the more unspecific, or normal tissue-like appearance of asymmetries in contrast to masses often appearing as more defined lesions. It seems reasonable that greater agreement regarding positive interpretation among radiologists applies to masses rather than asymmetries. This was also demonstrated in a study by Coolen et al, in which a higher proportion of masses and a lower proportion of asymmetries was observed among concordant compared with discordant recalls [13]. Another perspective is the radiological-pathological correlation of biopsies; a negative biopsy result is more likely to be considered representative in a benign-appearing asymmetry compared to a spiculated mass. Thus, some recalled lesions later diagnosed as interval cancer or next-round screen-detected cancer might have been biopsied with a negative result at prior screening. Reasons for a false negative biopsy include imprecise sampling at biopsy, or very subtle pathological findings and/or misinterpretation by the pathologists.

Our findings of no association between interpretation and histopathological tumor characteristics may indicate these characteristics to be of minor influence on the radiological detectability of the tumors. This is also in line with previous findings of the reviewed cases; a limited number of differences were observed between histopathological tumor characteristics, mammographic findings and review classification categories [10, 11].

Strengths and limitations

A major strength in our study is high data completeness and detailed information about the radiologists' interpretation scores obtained from the well-organized Cancer Registry of Norway. Further, the radiological review is one of the largest review studies reported on screen-detected and interval cancer including digital images only, and further, all analyses were breast-based.

Obviously, one major confounding factor applies to the findings in missed cancer; the presence of mammographic findings at prior screening influences both the frequency of positive interpretation as well as the classification of the cancer as missed. Further, our review design with a fully informed retrospective review with all images from diagnosis and prior screening available is associated with the highest proportion of missed cancers; in contrast to a blinded review in which the proportion of missed is shown to be substantially lower [8, 21, 22, 32].

Conclusions

The majority of mammograms with cancers were interpreted negatively by both screen-readers at the screening examination prior to diagnosis, including cancers classified as missed after an informed review of prior screening mammograms. However, about one third of the missed cancers had a positive interpretation score at prior screening, and the risk for a cancer to be missed was higher after a positive versus a negative interpretation at prior screening.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] N. Perry, M.J. Broeders, C. de Wolf, S. Tornberg, R. Holland, L. von Karsa, European guidelines for quality assurance in breast cancer screening and diagnosis, European Communities, Brussels, Belgium, 2006.
[2] IARC, Breast Cancer Screening IARC Handbook of Cancer Prevention Volume 15., Lyon, France, 2016.
[3] E.G. Klompenhouwer, A.C. Voogd, G.J. den Heeten, L.J. Strobbe, V.C. Tjan-Heijnen, M.J. Broeders, L.E. Duijm, Discrepant screening mammography assessments at blinded and non-blinded double reading: impact of arbitration by a third reader on screening outcome, Eur Radiol 25(10) (2015) 2821-9.
[4] S. Taylor-Phillips, C. Stinton, Double reading in breast cancer screening: considerations for policy-making, Br J Radiol 93(1106) (2020) 20190610.
[5] N. Perry, M. Broeders, C. de Wolf, S. Tornberg, R. Holland, L. von Karsa, European guidelines for quality assurance in breast cancer screening and diagnosis, European Communities, Brussels, Belgium, 2006.
[6] ECIBC, Recommendations from the European Breast Cancer Guidelines. https://healthcare-quality.jrc.ec.europa.eu/european-breast-cancer-guidelines Accessed June 2020. (Accessed June 1 2020).
[7] P.C. Brennan, A. Ganesan, M.P. Eckstein, E.U. Ekpo, K. Tapia, C. Mello-Thoms, S. Lewis, M.Z. Juni, Benefits of Independent Double Reading in Digital Mammography: A Theoretical Evaluation of All Possible Pairing Methodologies, Acad Radiol 26(6) (2019) 717-723.
[8] N. Houssami, L. Irwig, S. Ciatto, Radiological surveillance of interval breast cancers in screening programmes, Lancet Oncol 7(3) (2006) 259-65.
[9] N. Houssami, K. Hunter, The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening, NPJ Breast Cancer 3 (2017) 12.
[10] T. Hovda, S.R. Hoff, M. Larsen, L. Romundstad, K.K. Sahlberg, S. Hofvind, True and Missed Interval Cancer in Organized Mammographic Screening: A Retrospective Review Study of Diagnostic and Prior Screening Mammograms, Acad Radiol (2021).
[11] T. Hovda, K. Tsuruda, S.R. Hoff, K.K. Sahlberg, S. Hofvind, Radiological review of prior screening mammograms of screen-detected breast cancer, Eur Radiol 31(4) (2021) 2568-2579.
[12] M.A. Martiniussen, S. Sagstad, M. Larsen, A.S.F. Larsen, T. Hovda, C.I. Lee, S. Hofvind, Screen-detected and interval breast cancer after concordant and discordant interpretations in a population based screening program using independent double reading, Eur Radiol (2022).
[13] A.M.P. Coolen, J.R.C. Lameijer, A.C. Voogd, M.W.J. Louwman, L.J. Strobbe, V.C.G. Tjan-Heijnen, L.E.M. Duijm, Characteristics of screen-detected cancers following concordant or discordant recalls at blinded double reading in biennial digital screening mammography, Eur Radiol 29(1) (2019) 337-344.
[14] Ministry of Health and Care Services, Forskrift om innsamling og behandling av helseopplysninger i Kreftregisteret (The Cancer Registry Regulation), 2001. https://lovdata.no/dokument/SF/forskrift/2001-12-21-1477. (Accessed 01/12/2022.
[15] E.W. Bjørnson, A.S. Holen, S. Sagstad, M. Larsen, J. Thy, G. Mangerud, A.K. Ertzaas, S. Hofvind, BreastScreen Norway: 25 years of organized screening, 2022. https://www.kreftregisteret.no/Generelt/Rapporter/Mammografiprogrammet/25-arsrapport-mammografiprogrammet/. (Accessed 01/12/2022.
[16] Sickles E, D'Orsi CJ, Bassett LW, et al., ACR BI-RADS® Mammography. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System, American College of Radiology, Reston, VA, , 2013.
[17] S. Hofvind, B.M. Geller, R.D. Rosenberg, P. Skaane, Screening-detected breast cancers: discordant independent double reading in a population-based screening program, Radiology 253(3) (2009) 652-60.
[18] M. Roman, X. Castells, S. Hofvind, M. von Euler-Chelpin, Risk of breast cancer after false-positive results in mammographic screening, Cancer Med 5(6) (2016) 1298-306.

[19] M. von Euler-Chelpin, L.M. Risor, B.L. Thorsted, I. Vejborg, Risk of breast cancer after false-positive test results in screening mammography, J Natl Cancer Inst 104(9) (2012) 682-9.
[20] S. Hofvind, S. Sagstad, S. Sebuodegard, Y. Chen, M. Roman, C.I. Lee, Interval Breast Cancer Rates and Histopathologic Tumor Characteristics after False-Positive Findings at Mammography in a Population-based Screening Program, Radiology 287(1) (2018) 58-67.
[21] S. Hofvind, P. Skaane, B. Vitak, H. Wang, S. Thoresen, L. Eriksen, H. Bjorndal, A. Braaten, N. Bjurstam, Influence of review design on percentages of missed interval breast cancers: retrospective study of interval cancers in a population-based screening program, Radiology 237(2) (2005) 437-43.
[22] S. Ciatto, S. Catarzi, M.P. Lamberini, G. Risso, G. Saguatti, T. Abbattista, F. Martinelli, N. Houssami, Interval breast cancers in screening: the effect of mammography review method on classification, Breast 16(6) (2007) 646-52.
[23] E.U. Ekpo, M. Alakhras, P. Brennan, Errors in Mammography Cannot be Solved Through Technology Alone, Asian Pac J Cancer Prev 19(2) (2018) 291-301.
[24] J.G. Elmore, S.L. Jackson, L. Abraham, D.L. Miglioretti, P.A. Carney, B.M. Geller, B.C. Yankaskas, K. Kerlikowske, T. Onega, R.D. Rosenberg, E.A. Sickles, D.S. Buist, Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy, Radiology 253(3) (2009) 641-51.
[25] S. Hofvind, R.L. Bennett, J. Brisson, W. Lee, E. Pelletier, A. Flugelman, B. Geller, Audit feedback on reading performance of screening mammograms: An international comparison, J Med Screen 23(3) (2016) 150-9.
[26] T.D. Geertse, E. Paap, D. van der Waal, L.E.M. Duijm, R.M. Pijnappel, M.J.M. Broeders, Utility of Supplemental Training to Improve Radiologist Performance in Breast Cancer Screening: A Literature Review, J Am Coll Radiol 16(11) (2019) 1528-1546.
[27] P.D.Y. Trieu, K. Tapia, H. Frazer, W. Lee, P. Brennan, Improvement of Cancer Detection on Mammograms via BREAST Test Sets, Acad Radiol 26(12) (2019) e341-e347.
[28] A. Rodriguez-Ruiz, K. Lang, A. Gubern-Merida, M. Broeders, G. Gennaro, P. Clauser, T.H. Helbich, M. Chevalier, T. Tan, T. Mertelmeier, M.G. Wallis, I. Andersson, S. Zackrisson, R.M. Mann, I. Sechopoulos, Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists, J Natl Cancer Inst 111(9) (2019) 916-922.
[29] K. Lang, S. Hofvind, A. Rodriguez-Ruiz, I. Andersson, Can artificial intelligence reduce the interval cancer rate in mammography screening?, Eur Radiol (2021).
[30] I. Sechopoulos, J. Teuwen, R. Mann, Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art, Semin Cancer Biol 72 (2021) 214-225.
[31] A. Wadhwa, J.R. Sullivan, M.B. Gonyo, Missed Breast Cancer: What Can We Learn?, Curr Probl Diagn Radiol 45(6) (2016) 402-419.
[32] P.B. Gordon, M.J. Borugian, L.J. Warren Burhenne, A true screening environment for review of interval breast cancers: pilot study to reduce bias, Radiology 245(2) (2007) 411-5.

Figure legends

Figure 1. Study sample and review procedure for a retrospective, consensus-based, fully informed review of diagnostic and prior screening mammograms from 1223 women with screen-detected cancer and 1007 women with interval cancer.

Figure 2. Mammographic features (percentages) by interpretation at screening prior to diagnosis for missed screen-detected and interval cancer. Concordant negative: interpretation score 1 by both readers. Discordant: interpretation score 1 by one reader and ≥2 by the other. Concordant positive: Interpretation score ≥2 by both readers.

Tables

Interpretation at prior screening for the breast with later diagnosed cancer	Total	Missed	True negative	p-value
Screen-detected cancer				<0.001
Concordant negative	1009 (83%)	316 (69%)	693 (91%)	
Discordant	146 (12%)	92 (20%)	54 (7%)	
Concordant positive	68 (6%)	49 (11%)	19 (3%)	
Interval cancer				<0.001
Concordant negative	875 (87%)	312 (79%)	563 (92%)	
Discordant	111 (11%)	68 (17%)	43 (7%)	
Concordant positive	21 (2%)	16 (4%)	5 (1%)	
Interpretation at prior screening for the contralateral breast	**Total**	**Missed**	**True negative**	**p-value**
Screen-detected cancer				0.42
Concordant negative	1166 (95%)	433 (95%)	733 (96%)	
Discordant	35 (3%)	19 (4%)	16 (2%)	
Concordant positive	22 (2%)	5 (1%)	17 (2%)	
Interval cancer				0.20
Concordant negative	950 (94%)	380 (96%)	570 (93%)	
Discordant	47 (5%)	13 (3%)	34 (6%)	
Concordant positive	10 (1%)	3 (1%)	7 (1%)	

Table 1. Interpretation of the breasts with cancer and the contralateral breasts at screening prior to diagnosis for missed and true negative screen-detected and interval cancers classified in an informed retrospective review.

Concordant negative: interpretation score 1 by both readers. Discordant: interpretation score 1 by one reader and ≥2 by the other. Concordant positive: Interpretation score ≥2 by both readers.

Data given as numbers (percentages).

Interpretation at prior screening of the breast with cancer	OR (95% CI)
Screen-detected cancer	
Concordant negative	Reference
Discordant	3.8 (2.6-5.4)
Concordant positive	5.5 (3.2-9.5)
Interval cancer	
Concordant negative	Reference
Discordant	3.0 (2.0-4.5)
Concordant positive	6.3 (2.3-17.5)

Table 2. Age-adjusted odds ratio (OR) with 95% confidence interval (CI) for missed screen-detected and interval cancer for discordant and concordant positive interpretation with concordant negative as reference. Concordant negative: interpretation score 1 by both readers. Discordant: interpretation score 1 by one reader and ≥2 by the other. Concordant positive: Interpretation score ≥2 by both readers.

Positive interpretation score at prior screening for the breast later diagnosed with cancer	Total	Missed	True negative	p-value
Screen-detected cancer				0.10
Interpretation score 2	164 (77%)	102 (72%)	62 (85%)	
Interpretation score 3	46 (21%)	36 (26%)	10 (14%)	
Interpretation score 4	4 (2%)	3 (2%)	1 (1%)	
Interpretation score 5	.	.	.	
Interval cancer				0.68
Interpretation score 2	113 (86%)	70 (83%)	43 (90%)	
Interpretation score 3	15 (11%)	10 (12%)	5 (10%)	
Interpretation score 4	3 (2%)	3 (4%)	.	
Interpretation score 5	1 (1%)	1 (1%)	.	

Table 3. Highest positive interpretation score (2-5) for the breast with cancer at screening prior to diagnosis for missed and true negative screen-detected and interval cancer.
Data given as numbers (percentages).

Review classification category	Total recall	Recall due to findings in the breast later diagnosed with cancer	p-value
Screen-detected cancer			<0.001
Missed	74/457 (16.2%)	68/457 (14.9%)	
True negative	52/766 (6.8%)	32/766 (4.2%)	
Total	126/1223 (10.3%)	100/1223 (8.2%)	
Interval cancer			0.18
Missed	19/396 (4.8%)	16/396 (4.0%)	
True negative	23/611 (3.8%)	14/611 (2.3%)	
Total	42/1007 (4.2%)	30/1007 (3.0%)	

Table 4. Recall at screening prior to diagnosis. Total recall and recall due to findings in breast later diagnosed with cancer for cancers classfied as missed and true negative screen-detected and interval cancer.
Data given as numbers (percentages).

Histopathological characteristics	Concordant negative interpretation at prior screening	Discordant interpretation at prior screening	Concordant positive interpretation at prior screening
Screen-detected cancer	*n=1009*	*n=146*	*n=68*
Ductal carcinoma in situ	145 (14%)	23 (16%)	11 (16%)
Invasive carcinoma	864 (86%)	123 (84%)	57 (84%)
Invasive cancer only			
Median tumor diameter	13 mm (IQR 9-19)	13 mm (IQR 9-19)	12 mm (IQR 9-17)
Data not available	28	2	2
Histological grade 1	231 (27%)	34 (28%)	22 (42%)
Histological grade 2	431 (51%)	66 (55%)	22 (42%)
Histological grade 3	192 (23%)	21 (17%)	9 (17%)
Data not available	10	2	4
Lymph node positive	165 (20%)	26 (22%)	10 (18%)
Data not available	41	4	2
Estrogen receptor positive	756 (91%)	110 (92%)	52 (93%)
Data not available	30	4	1
Progesterone receptor positive	604 (73%)	85 (73%)	41 (73%)
Data not available	31	7	1
Interval cancer	*n=875*	*n=111*	*n=21*
Ductal carcinoma in situ	41 (5%)	4 (4%)	4 (19%)
Invasive carcinoma	834 (95%)	107 (96%)	17 (81%)
Invasive cancer only			
Median tumor diameter	19 mm (IQR 13-25)	18 mm (IQR 12-25)	16 mm (IQR 10-22)
Data not available	113	15	.
Histological grade 1	119 (15%)	7 (7%)	2 (12%)
Histological grade 2	369 (46%)	58 (56%)	9 (53%)
Histological grade 3	317 (39%)	39 (38%)	6 (35%)
Data not available	29	3	.
Lymph node positive	328 (41%)	51 (50%)	5 (29%)
Data not available	28	5	.
Estrogen receptor positive	648 (79%)	80 (78%)	15 (88%)
Data not available	16	5	.
Progesterone receptor positive	471 (58%)	58 (57%)	12 (71%)
Data not available	27	6	.

Table 5. Histopathological tumor characteristics for screen-detected and interval cancers with concorant negative, discordant or concordant interpretation at screening prior to diagnosis.
Concordant negative: interpretation score 1 by both readers. Discordant: interpretation score 1 by one reader and ≥2 by the other. Concordant positive: Interpretation score ≥2 by both readers.
Unless otherwise specified, data given as numbers (percentages).