PERFORMANCE OF DIAGNOSTIC MAMMOGRAPHY DIFFERS IN THE UNITED STATES AND DENMARK

Allan Jensen[1,2], **Berta M. Geller**[3], **Charlotte C. Gard**[4,5], **Diana L. Miglioretti**[4,5], **Bonnie Yankaskas**[6], **Patricia A. Carney**[7,8], **Robert D. Rosenberg**[9], **Ilse Vejborg**[10], and **Elsebeth Lynge**[1]

[1]Institute of Public Health, University of Copenhagen, Øster Farimagsgade 5, PO box 2099, DK-1014 Copenhagen K, Denmark

[2]Institute of Cancer Epidemiology, Danish Cancer Society, Strandboulevarden 49, DK-2100 Copenhagen O, Denmark

[3]University of Vermont, College of Medicine, Departments of Family Medicine and Radiology, Burlington, VT 05401-3444, US

[4]Group Health Research Institute, Biostatistics Unit, Seattle, WA 98101, US

[5]University of Washington, Department of Biostatistics, Seattle, WA, 98109, US

[6]University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7510, US

[7]Norris Cotton Cancer Centre, Dartmouth-Hitchcock Medical Centre, Department of Community and Family Medicine, Dartmouth Medical School, Lebanon, NH, US

[8]Oregon Health and Science University, School of Medicine, Department of Family Medicine, Portland, OR, US

[9]Department of Radiology, University of New Mexico - HSC, Albuquerque, NM 87131, US

[10]Centre of Diagnostic Imaging, University Hospital Copenhagen, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark

Abstract

Diagnostic mammography is the primary imaging modality to diagnose breast cancer. However, few studies have evaluated variability in diagnostic mammography performance in communities, and none has done so between countries. We compared diagnostic mammography performance in community-based settings in the United States and Denmark. The performance of 93,585 diagnostic mammograms from 180 facilities contributing data to the U.S. Breast Cancer Surveillance Consortium (BCSC) from 1999 through 2001 was compared to that of all 51,313 diagnostic mammograms performed at Danish clinics in 2000. We used the imaging workup's final assessment to determine sensitivity, specificity, and an estimate of accuracy: area under the receiver-operating characteristics (ROC) curve (AUC). Diagnostic mammography had slightly higher sensitivity in the United States (85%) than in Denmark (82%). In contrast, it had higher specificity in Denmark (99%) than in the United States (93%). The AUC was high in both countries: U.S. 0.91; and Denmark 0.95. Denmark's higher accuracy may result from supplementary ultrasound examinations, which are provided to 74% of Danish women but only 37% to 52% of U.S. women. In addition, Danish mammography facilities specialize in either

diagnosis or screening, possibly leading to greater diagnostic mammography expertise in facilities dedicated to symptomatic patients. Performance of community-based diagnostic mammography settings varied markedly between the two countries, indicating that it can be further optimized.

Keywords

Diagnostic mammography; comparative health care; breast cancer; breast neoplasms

INTRODUCTION

Unlike screening mammography, diagnostic mammography is used to identify breast cancers in women who have a breast symptom such as a lump. Diagnostic mammography is most often undertaken in combination with a clinical examination and supplemental breast ultrasound or other imaging (1). In all countries with mammography screening, the prevalence of breast cancer is approximately 10-fold higher and the stage of disease is more advanced in women receiving diagnostic mammography than in those receiving screening mammography (2).

Only a limited number of studies have examined the interpretive performance (i.e., sensitivity and specificity) of diagnostic mammography. Four studies examined the performance of diagnostic mammography in single facilities (3–6), whereas five additional studies have evaluated performance at the community level: four from the United States using data collected by the Breast Cancer Surveillance Consortium (BCSC) (2;7–10) and one from Denmark (11). The provision of diagnostic mammography differs between the United States and Denmark. In the United States, most mammography facilities offer both screening and diagnostic mammography and distinguish between the two services (12). By contrast, the organized mammography screening programmes in Denmark are run by clinics dedicated to screening, while women with breast symptoms are examined in clinics dedicated to diagnostic mammography (11).

In the United States, facilities participating in the BCSC, a population-based consortium of community mammography registries (13), prospectively collect data on all diagnostic mammograms, which are pooled at a statistical coordinating centre. Denmark has a complete database of diagnostic mammograms for 2000 (11).

Using these databases, the objective of this study was to compare the performance of diagnostic mammography across two community-based settings: in Denmark and in the United States. Our study is the first to evaluate variability in diagnostic performance between countries with differing systems of delivering mammography services. Our results suggest that distinctions between the diagnostic mammography programs in these countries may explain the variability in performance between them and suggest ways to deliver health services more effectively.

MATERIAL AND METHODS

In the United States, a diagnostic mammogram may include standard medio-lateral-oblique (MLO) and cranio-caudal (CC) projections, tangential views, or other special views to evaluate an area of clinical or radiographic concern such as spot compression or spot compression with magnification. When selecting a view, the proximity of the area of concern to the image receptor is considered (14). Breast ultrasound is also indicated in the "evaluation and characterization of palpable masses and other breast-related signs and/or symptoms" (15). Fine-needle aspiration (FNA) of the area of clinical concern, or to rule out simple cysts found at ultrasound, is not used routinely in the United States; however, some

radiologists use FNA instead of ultrasound to rule out cysts (16). In the United States, some symptomatic women also receive clinical breast examination (CBE), but it is not universal.

In Denmark, examination of women with breast concerns or symptoms includes obtaining a clinical history, CBE, diagnostic mammography using MLO and CC projections often supplemented with a medio-lateral (ML) projection, magnification, or spot compression projections. Indications for ultrasound are broader in Denmark than in the United States. In most clinics in Denmark, whole-breast ultrasound scanning is used for all palpable masses and all mammographic "probably benign" findings, suspicious abnormalities, and abnormalities highly suggestive of cancer. In case of suspicious abnormalities and abnormalities highly suggestive of cancer, the contra-lateral breast is also scanned with ultrasound. In Denmark, FNA or core needle biopsy (CNB) is used to investigate women presenting with all types of solid palpable breast lesions and also for non-palpable breast lesions with uncertain, suspicious, or malignant features on the diagnostic mammography or ultrasound examination (11;17;18).

We defined a diagnostic mammography as an examination of one woman, including at least one mammographic exposure performed to evaluate a breast symptom or clinical concern. We did not include diagnostic mammography performed for the additional evaluation or short-interval follow-up of a routine screening mammogram. If more than one diagnostic mammography was performed during the study period, we used the result of only the last examination to evaluate the performance. We included examinations performed with ultrasound or magnetic resonance imaging (MRI) (for both populations, MRI was rarely used around 2000) only if they were used as part of the additional workup of a diagnostic mammogram - not if they were the only imaging modality performed.

Data Sources

The United States—Data on diagnostic mammography in the United States were obtained from seven mammography registries that form the BCSC (13): 1) Carolina Mammography Registry, North Carolina 2) Group Health Cooperative, Seattle, Washington, 3) New Hampshire Mammography Network, 4) San Francisco Mammography Registry, California 5) Vermont Breast Cancer Surveillance System, 6) Colorado Mammography Project, Denver, and 7) New Mexico Mammography Project, Albuquerque. The BCSC uniformly collects data pertaining to mammography performance across diverse settings and populations (19). Data include patient demographic and clinical information, mammogram interpretation, biopsy results, and cancer diagnoses in the defined catchment areas of the participating facilities. Mammography results are reported using the categories of the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS®) (20). Cancer cases are ascertained through active follow-up and through linkages with state tumour registries or Surveillance, Epidemiology, and End Result (SEER) programs. Most registries supplement the capture of cancer cases through linkages with pathology databases. Cancer ascertainment from the tumour and SEER registries has been found to be at least 94% complete (2). We included diagnostic mammograms performed from January 1, 1999 through December 31, 2001, on women aged 20–89 without a history of breast cancer at 180 BCSC facilities. Each mammography registry and the statistical coordinating center have received institutional review board approval for either active or passive consenting processes or a waiver of consent to enroll participants, link data, and perform analytic studies. All procedures comply with the Health Insurance Portability and Accountability Act, and all registries and the statistical coordinating center have received a Federal Certificate of Confidentiality and other protection for the identities of women, physicians, and facilities that are subjects of this research.

Denmark—In Denmark, diagnostic mammography is available free to all Danish citizens after referral from a general practitioner. We collected data on all diagnostic mammograms performed in 2000 at all 32 public clinics, 12 private clinics, and 3 private hospitals in Denmark (11). Data were merged into a single database. Each record included the residence region, clinic, laterality of examined breast, date of examination, type of imaging and diagnosis, and the woman's unique personal identification number, which is used for all registration in Denmark. We used the same inclusion criteria for diagnostic mammograms in Denmark as for those in the United States. The Danish Data Protection Agency gave permission to undertake the study.

Classification of examinations as positive or negative for analysis

The United States—We used the final recorded assessment at the end of all imaging workup and within 3 months of the initial diagnostic mammography to evaluate the performance of a sequence of imaging that began with the diagnostic mammogram. Most of the examination records from the BCSC registries stated whether additional mammographic imaging or ultrasound was performed on the same day as the diagnostic mammographic examination and whether the result was used to arrive at the BI-RADS assessment; however, some registries did not completely capture whether an ultrasound was performed. Mammography examinations given a final BI-RADS assessment of 4 (suspicious abnormality) or 5 (highly suggestive of cancer) at the end of the imaging workup were considered positive (20). Examinations with a BI-RADS assessment of 3 (probably benign) or 0 (need additional imaging evaluation) and with a recommendation for biopsy, FNA, or surgical consultation were re-coded to BI-RADS code 4 (N=1, 354, 1.4 %). We classified as negative those mammograms given a final BI-RADS assessment of 1 (negative), 2 (benign), or 3 (probably benign, without a recommendation for biopsy, FNA, or surgical consultation) (20). Mammography examinations with a final BI-RADS assessment of 0 with a recommendation for additional imaging, non-specified workup, or missing recommendation were considered to be missing the final assessment, and thus were excluded from the analysis (N=1, 807, 1.9 %).

Denmark—Danish radiologists report on the outcome of a diagnostic mammography in free text. We collected these radiological reports for all Danish diagnostic mammograms in our database and scored the imaging outcome (combined evaluation of mammography and ultrasound, if performed) into 5 categories similar to the categorisation used in the BI-RADS classification system (11;20). The final assessments were based on the descriptors used in the reports (10). Normal reports were automatically scored 1; clear benign findings, 2; clear malignant reports, 5; and the remaining reports that were not obviously benign or malignant were scored into categories 3, 4, and 5 by three experienced radiologists (Ilse Vejborg and 2 experienced colleagues) who were blinded to the final diagnosis (cancer or no cancer). Mammography examinations scored as a final BI-RADS assessment of 4 or 5 at the end of the imaging workup were considered positive. We considered mammography examinations scored with a final assessment of 1, 2, or 3 as negative. As with the BCSC data, we used the final recorded assessment to evaluate the performance of a sequence of imaging that began with the diagnostic mammogram.

As a sensitivity analysis, we varied our definition of positive and negative in several ways. First, we considered all Danish mammograms with a BI-RADS 3 to be positive, given the proportion of cancer in these mammograms resembles that of mammograms given a BI-RADS 4 assessment in the United States. Second, we considered U.S. mammograms with a BI-RADS assessment of 0 or 3 with a recommendation for biopsy, FNA, or surgical consultation to be negative instead of positive.

Breast cancer status

To determine subsequent breast cancer status for U.S. women, the cohort of examined women was linked to regional cancer registries and pathology databases to determine cancer outcomes. Subsequent breast cancer status for Danish women was found by linkage with the Danish Cancer Register and the Danish Registry of Pathology, using the unique Danish personal identification numbers. We defined incident breast cancer as all invasive breast carcinoma or ductal carcinoma in situ (DCIS) that occurred within 1 year (365 days) of the initial diagnostic mammogram.

Performance measures and statistical analysis

For each woman, the BI-RADS assessment score and breast cancer status after 1 year were combined to identify true-positive (positive BI-RADS assessment with breast cancer), true-negative (negative BI-RADS assessment without a breast cancer), false-positive (positive BI-RADS assessment without breast cancer) and false-negative (negative BI-RADS with breast cancer) diagnostic mammography examinations. On this basis, sensitivity was defined as the percentage of positive examinations among women diagnosed with breast cancer [true-positive/(true-positive + false-negative)], and specificity was defined as the percentage of negative examinations among women without a breast cancer diagnosis [true-negative/ (true-negative + false-positive)]. We also used receiver operating characteristic (ROC) analysis to take into account the trade-off between sensitivity and specificity. For the ROC analysis, the modified BI-RADS assessment categories were ordered (1, 2, 3, 4, and 5) in accordance with an increasing likelihood of a breast cancer diagnosis. Accuracy was defined as the area under the receiver operating characteristic curve (AUC), with a value of 0.50 indicating purely random performance and 1.00 indicating the maximal value possible.

RESULTS

The study included 93,585 diagnostic mammograms from women at seven BCSC mammography registries (180 facilities) in the United States and 51,313 diagnostic mammograms from women at 47 clinics in Denmark (Table 1). In both the United States and Denmark, most diagnostic mammograms were performed in women aged 40–49 and 50–59 years. The proportion of U.S. women aged 40–49 years with a diagnostic mammography exceeded the proportion for the corresponding age group of Danish women (34% versus 27%); the reverse was true for women age 50–59 years (22% versus 29%). However, mean age at diagnostic mammogram was not markedly different between women in the two countries: 49.2 years in the United States and 49.8 years in Denmark. Based on BCSC registries with near complete capture of ultrasound, we estimated that ultrasound was used to make an assessment for 37% to 52% of all women. In contrast, ultrasound was used in 74% of examinations in Denmark (Table 1).

In the United States, 9,325 (10.0%) diagnostic mammograms were positive and 84 260 (90.0%) were negative In Denmark, 3,462 (6.7%) diagnostic mammograms were positive, and 47,851 (93.3%) were negative. The number of women with invasive breast cancer and DCIS diagnosed within 1 year of follow-up was 3,773 in the United States and 3,406 in Denmark, resulting in a lower proportion of women with breast cancer in the United States (4.0%) than in Denmark (6.6%). Despite the heavier burden of breast cancers in Denmark, a markedly higher proportion (91%) of the examined women in Denmark had normal/benign findings than did those in the United States (82%). Use of the BI-RADS categories "suspicious abnormality" and "highly suggestive of malignancy" somewhat compensated for this difference, because the proportion of women with these findings was lower in Denmark (6.7%) than in the United States (10.0%). However, the category "probably benign" which accounted for 8.0% of the examined women in the United States, and only for 2.5% in

Denmark, accounted for most of the difference. The proportion of women with cancer within each BI-RADS category also varied between the two countries, especially for BI-RADS codes 3, 4, and 5, where the proportions of women with breast cancer were markedly higher in Denmark than in the United States (Table 2).

The sensitivity of diagnostic mammography was higher in the United States data at 85.0% compared with 82.4% in the Danish data (Table 3). In contrast, the specificity was higher in Denmark with 98.6% compared with 93.2% in the United States. The area under the ROC curve (AUC) based on the 5-point BI-RADS assessment scale was higher in Denmark, 0.95, than in the United States, 0.91 (Table 3 and Figure 1).

If Danish mammograms given a BI-RADS 3 are considered positive, Denmark's sensitivity increases to 89.2%, higher than the U.S. sensitivity of 85.0%, and Denmark's specificity decreases to 96.5%, which is still higher than the U.S. specificity of 93.2%. Similarly, If U.S. mammograms with a BI-RADS assessment of 0 or 3 with a recommendation for biopsy, FNA, or surgical consultation are considered negative instead of positive, the U.S. sensitivity decreases to 81.6%, which is closer to the Danish sensitivity of 82.4%; however, the specificity only increases to 94.6%, which is still much lower than the Danish specificity of 98.6%.

DISCUSSION

Our study is the first to evaluate variability in performance of diagnostic mammography between the United States and a European country. This comparative study revealed different approaches in outcome coding and methods of workup between the two populations on performance of diagnostic mammography. Despite these reservations, the study suggests performance differences exist. Within 1 year of follow-up, the sensitivity was 2.6 percentage-points higher in the United States than in Denmark (85.0% compared to 82.4%), and the specificity was 5.4 percentage-points higher in Denmark than in the United States (98.6% compared to 93.2%). These differences in opposite directions suggest differences in the threshold used to recommend biopsy in the United States and Denmark; however, the overall AUC based on the 5-point BI-RADS assessment scale was higher in Denmark than in the United States, 0.95 versus 0.91, suggesting there is also a difference in accuracy. Accuracy may be higher in Denmark than in the United States due to supplementary ultrasound examinations provided to 74% of women in Denmark compared with only approximately 37% to 52% of women in the United States. In addition, Danish mammography facilities specialize in either diagnosis or screening, possibly leading to greater diagnostic mammography expertise in facilities dedicated to symptomatic patients.

One of the reasons for the observed differences in performance is that needle biopsy and ultrasound are sometimes used differently in Denmark and the United States for the evaluation of suspected fibroadenomas and cysts. A fibroadenoma usually has a characteristic appearance on ultrasound but not on mammography. In the United States, a palpable fibroadenoma-appearing mass will usually be classified as BI-RADS 4a (low suspicion of malignancy but warrants biopsy) or 4b (Intermediate suspicion of malignancy but warrants biopsy). This is a positive mammography, with a recommendation for a biopsy which decreases specificity. In the U.S. it is routine to biopsy what is believed to be a fibroadenoma on the small chance that it is a cancer while in Denmark during the time period of this study radiologists were willing to slightly lower their sensitivity to reduce the number of benign biopsies. Currently, however, Danish radiologists biopsy all solid palpable fibroadenomas.

Differences in coding may also account for some of the observed differences in performance. In the United States, diagnostic mammograms with a BI-RADS assessment of 0 (needs additional evaluation) or 3 (probably benign) with a recommendation for FNA, biopsy, or surgical consultation are considered positive, whereas in Denmark, all diagnostic mammograms with clearly benign signs were given BI-RADS codes 1–3, even if a biopsy was recommended. This results in a higher specificity in Denmark and a slightly lower sensitivity, given that some findings with benign features are cancer.

Another difference between the two countries is that some radiologists in the United States use FNA instead of ultrasound to rule our cysts. Mammograms with a BI-RADS assessment of 0 (needs additional evaluation) or 3 (probably benign) with a recommendation for FNA are considered positive exams. In Denmark, ultrasound examinations are used to rule out cysts, and these examinations are scored as negative. However, only 1.1% of U.S. mammograms had a final recommendation of FNA, so this could only account for a small proportion of the differences.

Data from BCSC included only mammograms individually coded as being performed for the evaluation of a clinical breast symptom or concern, whereas all mammograms from diagnostic mammography clinics in Denmark were included, allowing for a very minor contamination with opportunistic screening (21). For 2000, breast cancer incidence rates were 97 per 100,000 [World Health Organization (WHO) World Standard Population] white U.S. women, and 84 per 100,000 (World Standard Population) Danish women (22). Nevertheless, the Danish data set included a higher proportion of women with breast cancer. This may reflect that a high proportion of U.S. women had previously been screened, or that the criteria for referral differ: in 2000, 79.1 % of United States women aged 50+ reported having had a mammogram during the past two years (23), whereas only 28% of the Danish women aged 50+ were screened in the past two years before 2000.

Despite the perennial tradeoff between false- and true-positive examinations, diagnostic mammography aims to detect as many cancers with as few false-positive exams as possible. In the United States, for diagnostic mammograms, radiologists worry more about sensitivity than specificity, because women already have clinical symptoms and their breast cancers may be more advanced; so no one wants any additional delay in diagnosis. This contrasts with screening mammography, where the number of false-positives tests needs to be controlled, given the large number of women being screened, few of whom have cancer. The advantage of using ROC analysis in evaluating test accuracy compared with separate analyses of sensitivity and specificity is that the AUC provides an index of overall test accuracy if sensitivity and specificity have equal weights. However, computing the AUC on the 5-point BI-RADS scale to evaluate clinical performance has limitations (24–27).

Among 93,585 women included from the United States, 565 out of 3,773 breast cancers were missed at the diagnostic mammography, and 6,117 women without breast cancer were referred to biopsy. This shows that about 11 women had a false-positive test for each missed cancer. In contrast, when the Danish sensitivity and specificity is applied to the same number of women, 664 out of 3,773 breast cancers would have been missed, but only 1,227 women without breast cancer would have been referred to diagnostic workup. This shows that only two women will have a false-positive test for each missed cancer. Denmark needs to address the benefits and harms from missing cancers; and the United States needs to address the large number of biopsies performed in women without cancer.

Our study has several strengths. One is the degree of complete collection of standardized data for all diagnostic mammograms performed in Denmark in 2000 and for the BCSC. Furthermore, to stabilize our estimates, we included many women from both countries. The

BCSC data includes diverse areas of the United States and is representative of community practice in the United States. Our study also had some limitations. As with all comparisons between two health care delivery systems, we could not control subtle differences between them. In the United States, the BI-RADS codes were assessed by the clinical radiologist and guided the diagnostic process, while these codes were assessed retrospectively in Denmark based on the free text dictated at the reading session. Compared with the United States, Denmark had a more restrictive use of BI-RADS category 3 and a much larger proportion of breast cancers in this category. Because the BI-RADS codes were inferred retrospectively in Denmark based on the free texts, the codes might not fully reflect all management recommendations. In the United States, inconsistencies in the BI-RADS assessment and management recommendations for mammography are well documented (28;29). Access to mammography also differs between the two countries. Because diagnostic mammography is free and available in Denmark, women will not hesitate to seek care at the first sign of a symptom. We could not adjust for previous screening and time elapsed since last screen, which might have helped to explain some of the differences we found. We could not adjust for breast density, a risk factor for breast cancer that can affect accuracy; however, we have no reason to believe that the distributions vary between the two populations (30). We did not have complete capture of ultrasound examinations in the United States.

In conclusion, this is the first study to compare diagnostic mammography performance as it is practiced in community-based settings. Although the accuracy of diagnostic mammography was high in both countries, important differences exist between performance measures in the United States and in Denmark. Our findings therefore suggest that additional comparative studies will be useful to improve diagnostic service to women with signs and symptoms of breast cancer.

Acknowledgments

Abbreviations

AUC	area under the ROC curve
BCSC	Breast Cancer Surveillance Consortium
BI-RADS®	Breast Imaging Reporting and Data System
CBE	clinical breast examination
CC	cranio-caudal projection
DCIS	ductal carcinoma *in situ*
ML	medio-lateral projection
MLO	medio-lateral-oblique projection

MRI	magnetic resonance imaging
ROC	receiver operating characteristic
FNA	fine needle aspiration
CNB	core needle biopsy
WHO	World Health Organization

References

1. Perry NM. Quality assurance in the diagnosis of breast disease. EUSOMA Working Party. Eur J Cancer. 2001 Jan.37:159–72. [PubMed: 11166142]

2. Sickles EA, Miglioretti DL, Ballard-Barbash R, Geller BM, Leung JW, Rosenberg RD, Smith-Bindman R, Yankaskas BC. Performance benchmarks for diagnostic mammography. Radiology. 2005; 235:775–90. [PubMed: 15914475]

3. Flobbe K, van der Linden ES, Kessels AG, van Engelshoven JM. Diagnostic value of radiological breast imaging in a non-screening population. Int J Cancer. 2001; 92:616–8. [PubMed: 11304700]

4. Eltahir A, Jibril JA, Squair J, Heys SD, Ah-See AK, Needham G, Gilbert FJ, Deans HE, McKean ME, Smart LM, Eremin O. The accuracy of "one-stop" diagnosis for 1,110 patients presenting to a symptomatic breast clinic. J R Coll Surg Edinb. 1999; 44:226–30. [PubMed: 10453144]

5. Zonderland HM, Pope TL Jr, Nieborg AJ. The positive predictive value of the breast imaging reporting and data system (BI-RADS) as a method of quality assessment in breast imaging in a hospital population. Eur Radiol. 2004; 14:1743–50. [PubMed: 15243715]

6. Duijm LE, Guit GL, Zaat JO, Koomen AR, Willebrand D. Sensitivity, specificity and predictive values of breast imaging in the detection of cancer. Br J Cancer. 1997; 76:377–81. [PubMed: 9252206]

7. Barlow WE, Lehman CD, Zheng Y, Ballard-Barbash R, Yankaskas BC, Cutter GR, Carney PA, Geller BM, Rosenberg R, Kerlikowske K, Weaver DL, Taplin SH. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. J Natl Cancer Inst. 2002; 94:1151–9. [PubMed: 12165640]

8. Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. Radiology. 2002; 224:861–9. [PubMed: 12202726]

9. Miglioretti DL, Smith-Bindman R, Abraham L, Brenner RJ, Carney PA, Bowles EJ, Buist DS, Elmore JG. Radiologist characteristics associated with interpretive performance of diagnostic mammography. J Natl Cancer Inst. 2007; 99:1854–63. [PubMed: 18073379]

10. Jackson SL, Taplin SH, Sickles EA, Abraham L, Barlow WE, Carney PA, Geller B, Berns EA, Cutter GR, Elmore JG. Variability of interpretive accuracy among diagnostic mammography facilities. J Natl Cancer Inst. 2009; 101:814–27. [PubMed: 19470953]

11. Jensen A, Vejborg I, Severinsen N, Nielsen S, Rank F, Mikkelsen GJ, Hilden J, Vistisen D, Dyreborg U, Lynge E. Performance of clinical mammography: a nationwide study from Denmark. Int J Cancer. 2006; 119:183–91. [PubMed: 16450388]

12. National Cancer Institute DoCCaPSARP. Breast Cancer Surveillance Consortium. Evaluating Screening Performance in Practice. Bethesda, Maryland: 2004.

13. Ballard-Barbash R, Taplin SH, Yankaskas BC, Ernster VL, Rosenberg RD, Carney PA, Barlow WE, Geller BM, Kerlikowske K, Edwards BK, Lynch CF, Urban N, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol. 1997; 1694:1001–8. [PubMed: 9308451]

14. [Accessed 22 October 2009] 2009. http://www.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/breast.aspx

15. [Accessed 22 October 2009] 2009. http://www.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/breast/us_breast.asp

16. Kerlikowske K, Smith-Bindman R, Ljung BM, Grady D. Evaluation of abnormal mammography results and palpable breast abnormalities. Ann Intern Med. 2003; 139:274–84. [PubMed: 12965983]

17. National Board of Health. Guideline on diagnostic breast disease assessment [In Danish]. Copenhagen: 1999.

18. Jensen A, Rank F, Dyreborg U, Severinsen N, Nielsen S, Lynge E, Vejborg I. Performance of combined clinical mammography and needle biopsy: a nationwide study from Denmark. APMIS (Copenhagen). 2006; 114:884–92.

19. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, Geller BM, Abraham LA, Taplin SH, Dignan M, Cutter G, Ballard-Barbash R. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med. 2003; 138:168–75. [PubMed: 12558355]

20. American College of Radiology (ACR). Breast Imaging Reporting and Data System Atlas (BI-RADS® Atlas). VA: Reston; 2003.

21. Jensen A, Olsen AH, Euler-Chelpin M, Helle NS, Vejborg I, Lynge E. Do nonattenders in mammography screening programmes seek mammography elsewhere? Int J Cancer. 2004; 113:464–70. [PubMed: 15455383]

22. [Accessed 12 October 2009] 2009. http://www-dep.iarc.fr/CI5-IX/PDF/BYSITE/S_C50.pdf

23. [accessed Sept 5, 2009] 2009. http://apps.nccd.cdc.gov/BRFSS/list.asp?cat=WH&yr=2000&qkey=4427&state=All

24. Obuchowski NA. Receiver operating characteristic analysis: A proper measurement for performance in breast cancer screening?: Author reply. Am J Roentgenol. 2006; 186:580. [PubMed: 16423974]

25. Pepe MS. A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. Biometrika. 1997; 84:595–608.

26. Krupinski EA, Jiang Y. Anniversary paper: evaluation of medical imaging systems. Med Phys. 2008; 35:645–59. [PubMed: 18383686]

27. Hadjiiski L, Chan HP, Sahiner B, Helvie MA, Roubidoux MA. Quasi-continuous and discrete confidence rating scales for observer performance studies: Effects on ROC analysis. Acad Radiol. 2007; 14:38–48. [PubMed: 17178364]

28. Geller BM, Barlow WE, Ballard-Barbash R, Ernster VL, Yankaskas BC, Sickles EA, Carney PA, Dignan MB, Rosenberg RD, Urban N, Zheng Y, Taplin SH. Use of the American College of Radiology BI-RADS to report on the mammographic evaluation of women with signs and symptoms of breast disease. Radiology. 2002; 222:536–42. [PubMed: 11818625]

29. Geller BM, Ichikawa LE, Buist DS, Sickles EA, Carney PA, Yankaskas BC, Dignan M, Kerlikowske K, Yabroff KR, Barlow W, Rosenberg RD. Improving the concordance of mammography assessment and management recommendations. Radiology. 2006; 241:67–75. [PubMed: 16990672]

30. Olsen AH, Bihrmann K, Jensen MB, Vejborg I, Lynge E. Breast density and outcome of mammography screening: a cohort study. Br J Cancer. 2009; 100:1205–8. [PubMed: 19293800]
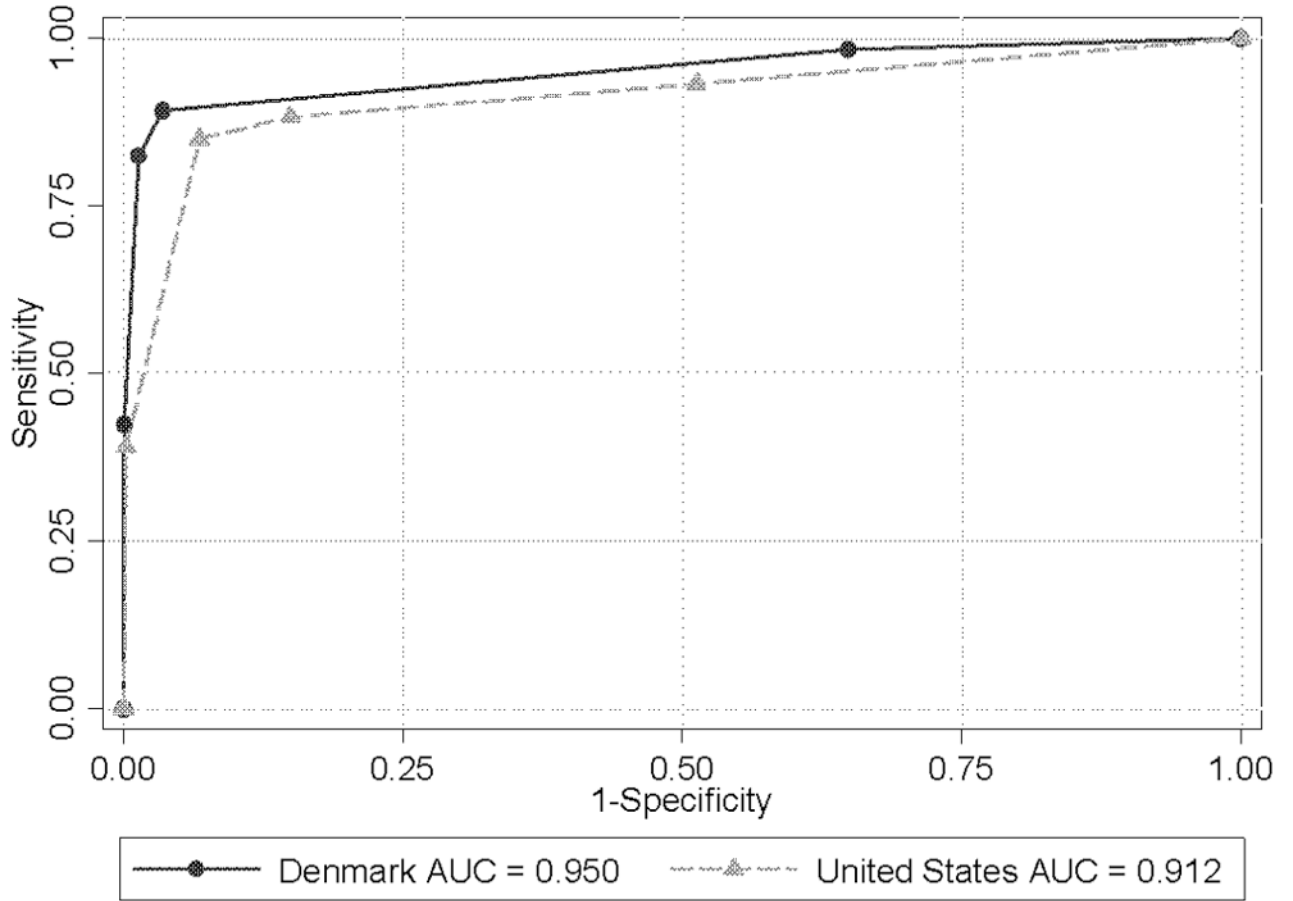
Figure 1.
Diagnostic mammography empirical receiver operating characteristic (ROC) curves for Denmark and the United States, respectively. Empirical or observed sensitivity is plotted against the empirical or observed false-positive rate (1 minus specificity) for each BI-RADS© (Breast Imaging Reporting and Data System) criterion point. Starting from the lower left, the cut points for the mammographic examination BI-RADS™ assessments being called positive are 5, 4, 3, 2, and 1. AUC is area under the ROC curve.

Table 1

Characteristics of study participants from United States (Breast Cancer Surveillance Consortium) and Denmark.

	United States	Denmark
Study base	180 breast imaging facilities from 7 population-based mammography registries of the BCSC	All 47 diagnostic mammography clinics in Denmark (screening clinics not included)
Definition of diagnostic mammogram	Mammograms recorded as evaluation of breast problem (symptomatic)	All mammograms from diagnostic clinics[1]
Study period	1999–2001	2000
Total number of women	93,585	51,313
Age distribution, years: number (%)[2]		
20–29	2997 (3.2)	1906 (3.7)
30–39	19 632 (21.0)	8710 (17.0)
40–49	31 421 (33.6)	13 582 (26.5)
50–59	20 147 (21.5)	14 871 (29.0)
60–69	10 510 (11.2)	7653 (14.9)
70–79	6725 (7.2)	3697 (7.2)
80–89	2153 (2.3)	894 (1.7)
Breast cancer definition	Histology-confirmed invasive carcinoma or DCIS	Histology/cytology-confirmed invasive/DCIS
Breast cancer identification	Linkage with cancer registries and pathology databases	Linkage with nationwide cancer and pathology registries
Examinations supplemented with ultrasound	37% to 52%[3]	74%
Breast cancer incidence[4]	97 (USA, white)	84 (Denmark)

[1] Outside the two organized screening programmes, asymptomatic women can be referred to diagnostic mammography clinics for opportunistic screening. This happened seldom, as the annual utilization rate of all diagnostic mammography (symptomatic and asymptomatic women together) was only 2.7% in women aged >25 in 2000 (21).

[2] Number of women (%)

[3] Depending on registry.

[4] Age-standardized rate per 100,000 (WHO World Standard Population) for the year 2000 (22).

Table 2

Number (%) of women with diagnostic mammography by BI-RADS code and number (%) with breast cancer within 1 year of follow-up in community-based studies of diagnostic mammography from United States and Denmark.

	BI-RADS code	Number of women (%*)		Number and proportion (%**) of women with breast cancer	
		United States	Denmark	United States	Denmark
Negative examinations	1	43 954 (47.0)	16 863 (32.9)	257 (0.6)	57 (0.3)
	2	32 857 (35.1)	29 728 (57.9)	183 (0.6)	310 (1.0)
	3	7449 (8.0)	1260 (2.5)	125 (1.7)	232 (18.4)
	Total	84 260 (90.0)	47 851 (93.3)	565 (0.7)	599 (1.3)
Positive examinations	4	7645 (8.1)	1973 (3.8)	1737 (22.7)	1362 (69.0)
	5	1680 (1.8)	1489 (2.9)	1471 (87.6)	1445 (97.0)
	Total	9325 (10.0)	3462 (6.7)	3208 (34.4)	2807 (81.1)
All examinations	Total	93 585 (100.0)	51 313 (100.0)	3773 (4.0)	3406 (6.6)

BI-RADS codes: 1: Negative; 2: Benign; 3: Probably benign; 4: Suspicious abnormality; 5: Highly suggestive of cancer.

*
Column percentages

**
Row percentages of number of women in country with BI-RADS code.

Table 3

Performance of diagnostic mammography in the United States and Denmark.

	United States	Denmark
True-negative tests (%)	83 695 (89.4)	47 252 (92.1)
True-positive tests (%)	3208 (3.4)	2807 (5.5)
False-negative tests (%)	565 (0.6)	599 (1.2)
False-positive tests (%)	6117 (6.5)	655 (1.3)
Sensitivity (%)	85.0	82.4
Specificity (%)	93.2	98.6
Accuracy (AUC)	0.912	0.950