Title
Assessment of Radiologist Performance in Breast Cancer Screening Using Digital Breast Tomosynthesis vs Digital Mammography

Journal
JAMA Network Open, 3(3)

ISSN
2574-3805

Authors
Sprague, Brian L
Coley, R Yates
Kerlikowske, Karla
et al.

Publication Date
2020-03-02

DOI
10.1001/jamanetworkopen.2020.1759

Peer reviewed

HHS Public Access
Author manuscript
JAMA Netw Open. Author manuscript; available in PMC 2021 March 02.

Published in final edited form as:
JAMA Netw Open. ; 3(3): e201759. doi:10.1001/jamanetworkopen.2020.1759.

Assessment of Radiologist Performance in Breast Cancer Screening Using Digital Breast Tomosynthesis vs Digital Mammography

Brian L. Sprague, PhD,
Departments of Surgery and Radiology, University of Vermont Cancer Center, University of Vermont, Burlington, VT, USA.

R Yates Coley, PhD,
Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA.

Karla Kerlikowske, MD, MS,
Departments of Medicine and Epidemiology and Biostatistics, University of California, San Francisco, CA, USA and General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco, CA, United States.

Garth H. Rauscher, PhD, MPH,
Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, IL, USA.

Louise M. Henderson, PhD,
Departments of Radiology and Epidemiology, University of North Carolina, Chapel Hill, NC, USA.

Tracy Onega, PhD,
Departments of Biomedical Data Science and Epidemiology, The Dartmouth Institute for Health Policy & Clinical Practice and Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA.

Christoph I. Lee, MD, MS,
Department of Radiology, University of Washington School of Medicine, and the Hutchinson Institute for Cancer Outcomes Research, Seattle, WA, USA.

Sally D. Herschorn, MD,

CORRESPONDENCE: Brian L. Sprague, PhD, 1 S. Prospect St, UHC Room 4425, University of Vermont, Burlington, VT 05401, (t) 802-656-4112; (f) 802-656-8826; bsprague@uvm.edu.

Author Contributions
Drs. Coley and Miglioretti had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Sprague, Coley, Kerlikowske, Rauscher, Miglioretti.
Acquisition, analysis, or interpretation of the data: All authors.
Drafting of the manuscript: Sprague
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Coley
Obtained funding: Sprague, Kerlikowske, Rauscher, Henderson, Onega, Lee, Tosteson, Miglioretti

Department of Radiology, University of Vermont Cancer Center, University of Vermont, Burlington, VT, USA.

Anna N.A. Tosteson, ScD,
The Dartmouth Institute for Health Policy and Clinical Practice and Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA.

Diana L. Miglioretti, PhD
Division of Biostatistics, Department of Public Health Sciences, University of California Davis School of Medicine, Davis, CA, USA and Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle, WA, USA.

Abstract

Importance: Many US radiologists have screening mammography recall rates above the expert-recommended threshold of 12%. The impact of digital breast tomosynthesis (DBT) on the distribution of radiologist recall rates is uncertain.

Objective: To evaluate radiologists' recall and cancer detection rates after vs. before beginning interpretation of DBT exams.

Design: Prospective observational cohort.

Setting: 104 radiology facilities in the Breast Cancer Surveillance Consortium.

Participants: 198 radiologists who interpreted 251,384 DBT and 2,000,681 digital mammography (DM) screening exams during 2009–2017, including 126 who interpreted DBT exams during the study period and 72 who exclusively interpreted DM exams (to adjust for secular trends).

Exposures: DBT and DM screening exams.

Main outcomes and measures: Recall rate, cancer detection rate.

Results: Among radiologists who interpreted DBT, 65.9% had unadjusted DM recall rates ≤12% prior to using DBT, with a median recall rate of 10.0% (25^{th}-75^{th} percentile=7.5–13.0%). On DBT exams, 76.2% had an unadjusted recall rate ≤12%, with a median recall rate of 8.8% (25^{th}-75^{th}=6.3–11.3%). A secular trend in recall rate was observed, with the multivariable-adjusted risk of recall on screening exams declining by 1.2% per year (95% CI: 0.9–1.5%). After adjusting for exam characteristics and secular trends, recall rates were 15% lower on DBT compared to DM exams interpreted prior to DBT use (relative risk [RR]=0.85, 95% confidence interval [CI]=0.83–0.87). Adjusted recall rates were significantly lower on DBT compared to DM exams interpreted prior to DBT use for 35.7% of radiologists and significantly higher for 14.3%; 50.0% had no statistically significant change. The unadjusted cancer detection rate on DBT was 5.3 per 1000 exams (95% CI: 5.0–5.7), compared to 4.7 per 1000 (95% CI: 4.6–4.8) on DM exams interpreted prior to DM use (multivariable-adjusted RR=1.21; 95% CI=1.11–1.33).

Conclusions and Relevance: DBT is associated with an overall decrease in recall rate and an increase in cancer detection rate. However, we demonstrate for the first time that there is wide variability across radiologists, including a subset of radiologists who experienced increased recall rates on DBT. Radiology practices should audit radiologist DBT screening performance and

consider additional DBT training for radiologists whose performance does not improve as expected.

INTRODUCTION

In the United States, high false-positive rates on mammography have been recognized as a significant harm of breast cancer screening.[1] Only 4–5% of positive mammograms recalled for further evaluation ultimately lead to a cancer diagnosis.[2] There have long been calls for quality improvement efforts to lower screening recall rates in the US,[3] but little evidence to date of improvements. The American College of Radiology professional guidelines for mammography interpretation[4] issued in 2013 included a recommended upper threshold of 12% for recall rate, citing the findings of a panel of expert breast imaging physicians.[5] In an evaluation of digital mammography (DM) screening performance in the Breast Cancer Surveillance Consortium (BCSC) during 2007–2013, only 62.2% of 359 radiologists had a recall rate below the expert-recommended upper threshold of 12%.[2] The median recall rate among BCSC radiologists increased from 9.7% on film-screen mammography during 1996–2002 to 10.8% on DM during 2007–2013.[6]

Digital breast tomosynthesis (DBT) is a new mammography-based tool for breast cancer screening that has quickly disseminated. DBT acquires multiple low-dose 2D mammography images that are reconstructed computationally and can be scrolled through in a 3D format.[7,8] This approach is designed to clarify areas of overlapping breast tissue that may obscure lesions. Approved by the Food and Drug Administration (FDA) in 2011, about half of FDA-certified mammography facilities now have a DBT-capable mammography unit.[9]

Prior studies suggest that DBT screening exams have on average lower recall rates in the US compared to conventional DM, while maintaining or even increasing cancer detection.[10] To our knowledge, radiologist-level variability in DBT performance has not been examined among a large national sample. We compared radiologist recall and cancer detection rates with DBT and DM using observational clinical data from the BCSC, which includes a large sample of radiologists from diverse practice settings in the US. We hypothesized that most radiologists would have a lower recall rate on DBT screens compared to their recall rate on DM screens interpreted before beginning the use of DBT, without an adverse impact on cancer detection.

METHODS

Study Setting and Data Sources

We used data from the six active BCSC registries (http://www.bcsc-research.org/index.html): Carolina Mammography Registry, Kaiser Permanente Washington, New Hampshire Mammography Network, Vermont Breast Cancer Surveillance System, San Francisco Mammography Registry, and Metropolitan Chicago Breast Cancer Registry.[2,11] BCSC registries capture exam-level risk factor and radiology data directly from participating radiology facilities. Data on breast cancer diagnoses are obtained by linking to pathology databases; regional Surveillance, Epidemiology, and End Results programs; and state tumor registries. Each registry and the Statistical Coordinating Center received institutional review board approval for either active or passive consenting processes or a waiver of consent. All procedures were Health Insurance Portability and Accountability Act compliant. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cohort studies.

Participants

Radiologists who interpreted screening mammography at participating BCSC facilities were eligible for the study. For each radiologist who had interpreted DBT exams, we determined the date of their first DBT exam interpretation (range: June 2011-November 2016). Three groups of screening mammograms were defined for each radiologist who initiated use of DBT during the study period: 1) DM exams in the 2 years before the radiologist's DBT start date; 2) DBT exams; and 3) DM exams after the radiologist's DBT start date. Radiologists who interpreted DBT were required to have interpreted at least 100 DBT exams and at least 100 DM exams prior to using DBT (N=126) to ensure reasonable precision[12] for the primary comparison of recall rate across groups while also maintaining representation of low volume radiologists.

To estimate and control for secular trends in mammography performance, we also identified radiologists who did not interpret DBT during the study period ("DBT non-users"). We restricted these to the 72 radiologists who read 960 or more DM exams (the minimum required by the Mammography Quality Standards Act) in the BCSC database during the two year period surrounding July 23, 2014 (the median DBT start date for DBT interpreters).

Measures and definitions

Demographic, risk factor, and medical history information for women undergoing breast imaging was obtained on a self-administered questionnaire completed at each mammogram or by extraction from the electronic medical record. We estimated each woman's 5-year breast cancer risk using the BCSC version 2.0 risk model.[13]

Breast imaging data including modality, exam indication, breast density, and assessments were provided by radiology facilities using standard nomenclature defined by the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS).[4] All analyses were restricted to DM and DBT screening exams conducted during 2009–2017 among women without a personal history of breast cancer or a history of breast augmentation, as screening performance is known to differ markedly among these women.[14,15] Recall rate was defined as the fraction of screening exams with a positive initial assessment (category 0, 3, 4, or 5) based on American College of Radiology guidelines.[4] Cancer detection rate was defined as the number of positive exams with invasive cancer or DCIS diagnosed within 365 days and prior to the next screening mammogram divided by the total number of exams.

Statistical Analyses

Descriptive statistics were used to evaluate the unadjusted recall rates and cancer detection rates within comparison groups. To control for secular trends and differences in covariates across comparison groups, a generalized linear model with log link and robust variance estimation was used to estimate the probability of a positive exam (i.e., recall) associated with four different comparison groups: DM interpreted prior to any DBT interpretation ("DM pre-DBT"), DBT exams, DM exams interpreted after DBT interpretation ("DM post-DBT"), and DM exams interpreted by radiologists who did not interpret DBT during the study period. Exam-level covariates in the models were identified *a priori* based on known association with screening performance, and included age, race/ethnicity, first degree family history of breast cancer, history of breast biopsy, BI-RADS breast density, BCSC 5-year risk, time since last mammogram, history of prior DBT imaging, exam year, and BCSC registry site. We included radiologist-level fixed effects to account for differences in performance due to radiologist-level factors associated with DBT use or performance; thus, the estimated relative risks are within-radiologists effects that are adjusted for between radiologists differences, exam-level covariates, and secular trends.[16] For radiologists that interpreted DBT during the study period, separate radiologist-level fixed effects were estimated for DM pre-DBT screens, DBT screens, and DM post-DBT screens. We estimated adjusted recall rates over time for each comparison group using model-based estimates of exam year effects and comparison-group-specific radiologist-level fixed effects, adjusted for the average exam characteristics.

Radiologist characteristics were described according to whether the radiologist interpreted DBT during the study period and, for DBT interpreters, whether a change in adjusted recall rate on DBT vs. DM was observed. Chi-square tests were used to compare radiologist characteristics among radiologist groups that had a statistically significant decrease, no significant change, and statistically significant increase in recall rate on DBT compared to pre-DBT DM screens.

Sample size was insufficient to evaluate radiologist-level cancer detection rates for each comparison group in our study. However, we estimated the overall relative risk of cancer detection by comparison group using a generalized linear regression model with log link, using a similar approach as described above for recall including radiologist-level fixed effects but excluding radiologist-level and comparison group interactions. We restricted these analyses to exams with at least 365 days of follow-up for cancer diagnoses (N=2,037,013). We estimated adjusted cancer detection rates over time for each comparison group using model-based estimates of exam year effects and radiologist-level fixed effects, adjusted for the average exam characteristics.

Statistical analyses were performed using R software (version 3.5; R Foundation for Statistical Computing, Vienna, Austria). Tests of statistical significance were two-sided with an alpha of 0.05.

RESULTS

Characteristics of women undergoing screening were generally similar across comparison groups (Table 1), though women undergoing DBT exams were more likely to be non-Hispanic white and less likely to be non-Hispanic black or Asian/Pacific Islander. Women with DM exams interpreted by radiologists who did not interpret DBT during the study period were more likely to have BCSC 5-year risk ≥1.67% compared to women with DM or

DBT exams interpreted by DBT users. The majority of radiologists were not breast imaging specialists and did not practice at an academic medical center (Table 2). Radiologists who interpreted DBT were more likely to be breast imaging specialists and fellowship trained, and more likely to practice at an academic medical center, in a hospital, and in a for-profit practice compared to radiologists who did not interpret DBT during the study period.

Recall Rate

DBT vs. DM before DBT use—Among radiologists who interpreted DBT, the median unadjusted DM recall rate before DBT use was 10.0%, and 65.9% of radiologists had a recall rate below the expert-recommended upper threshold of 12% (Table 3; Figure 1A). The median unadjusted DBT recall rate was 8.8%, with 76.2% of radiologists meeting the 12% threshold. After adjusting for exam-level covariates and secular trends, the overall relative risk for recall on DBT vs. DM pre-DBT exams was 0.85 (95% CI:0.83–0.87; Table 3).

There was a statistically significant decrease in the multivariable-adjusted recall rate for 45 radiologists (35.7%), a statistically significant increase for 18 radiologists (14.3%), and no statistically significant difference for 63 radiologists (50.0%) (Figure 1B). Radiologist characteristics were comparable across these groups in univariate tests, with the exception of the relative amount of DBT vs. DM interpreted in the "post-DBT" period (p=0.01; Table 2). The majority (60%) of radiologists who experienced a decrease in recall rate on DBT had at least 50% DBT volume (vs. DM) after beginning DBT interpretation.

DM before vs. after DBT use—While the unadjusted statistics also suggested a reduction in recall rate on DM exams interpreted after DBT use, the multivariable-adjusted model estimates indicated comparable recall rates to DM exams interpreted prior to DBT use (RR=1.01, 95% CI:0.99–1.03; Table 3).

DBT vs. DM among radiologists who did not interpret DBT—The median unadjusted DM recall rate among radiologists who did not interpret DBT was 9.5% (IQR: 6.5–10.5%), with 84.7% of radiologists meeting the 12% recall rate threshold (eTable 1, Supplemental Material). A secular trend was observed, with the multivariable-adjusted risk of recall on screening exams declining by 1.2% per year (95% CI:0.9–1.5%). Figure 2A illustrates the multivariable-adjusted model estimates of recall rate in each comparison group in the presence of this secular trend, and also demonstrates that after adjustment for secular trends and woman-level characteristics, recall rates on DBT were lower than recall rates on DM among radiologists who did not interpret DBT.

Cancer Detection Rate

DBT vs. DM before DBT use—Among radiologists who interpreted DBT, the unadjusted cancer detection rate on DM exams prior to DBT use was 4.7 per 1000 exams (95% CI:4.6–4.8) and the unadjusted cancer detection rate on DBT exams was 5.3 per 1000 (95% CI:5.0–5.7; p<0.0001) (Table 3). In the multivariable-adjusted regression analysis, the relative risk of cancer detection on DBT vs. DM exams prior to DBT use was 1.21 (95% CI:1.11–1.33; Table 3). Among radiologists who had a statistically significant decrease in recall rate on DBT compared to DM pre-DBT (N=45; median DBT recall rate 7.5%), the cancer detection rate on DBT exams was 5.0 per 1000 (95% CI:4.6–5.4) compared to 4.7 per 1000 (95% CI:4.5–4.9) on DM pre-DBT exams (p=0.22). Among radiologists who did not have a statistically significant decrease in recall on DBT (N=81; median DBT recall rate 9.1%) the cancer detection rate on DBT exams was 5.8 per 1000 (95% CI:5.2–6.3) compared to 4.7 per 1000 (95% CI:4.5–4.9) on DM pre-DBT exams (p<0.0001).

DM before vs. after DBT use—While the unadjusted statistics also suggested an increase in cancer detection rate on DM exams interpreted after DBT use, the multivariable-adjusted model estimates indicated comparable cancer detection rates to DM exams interpreted prior to DBT use (RR=1.06, 95% CI:0.98–1.14; Table 3).

DBT vs. DM among radiologists who did not interpret DBT—The unadjusted statistics indicated that radiologists who did not interpret DBT had relatively high cancer detection rates (eTable 1, Supplemental Material), but the multivariable-adjusted model indicated that after adjustment for secular trends and woman-level characteristics, cancer detection rates on DBT were higher than cancer detection rates on DM among radiologists who did not interpret DBT (Figure 2B). In the multivariable-adjusted regression model the secular trend in cancer detection rates was small and not statistically significant (annual change 0.4%, 95% CI:−1.8–1.0%).

DISCUSSION

Our results demonstrate a clinically important downward shift in the distribution of radiologist recall rates with DBT screening among a large geographically-diverse sample of US radiologists. This shift was accompanied by elevated cancer detection rates on DBT exams, indicating that the reduction in recall did not come at the sacrifice of cancer detection. Improvements in screening outcomes remained after adjusting for differences in woman- and exam-level characteristics and secular trends.

Overall, our results are consistent with prior studies evaluating average differences in recall rates on DBT vs. DM exams in the US.[10,17–19] A recent meta-analysis of thirteen US studies found that the overall DBT recall rate was 2.2% points lower than DM, while the cancer detection rate was elevated by 1.1 per 1000 exams.[10] However, our study demonstrates that the impact of DBT on recall rate varies widely across radiologists. The observed variability indicates that decreased recall on DBT is not universal. We found no strong differences in DBT vs. DM recall rate patterns in relation to radiologist specialty, fellowship training, or practice characteristics, but radiologists who shifted towards predominant use of DBT (>50% of screening volume) were more likely to have a decrease in recall rate on DBT compared to radiologists who continued interpreting a large percentage of DM screens after beginning DBT. Additional evidence is needed to characterize radiologist- and practice-level factors associated with DBT performance, and to identify settings where additional training or other interventions are warranted to realize the potential benefits of DBT.

Our analyses were designed to test the hypothesis that radiologist-specific recall rates on DBT were lower than their recall rates on DM prior to the use of DBT. We included a group of radiologists who did not interpret DBT to estimate and adjust for secular trends. The reasons for the declining secular trend in recall rate observed in our study are not clear, but we speculate that it may be due to increased attention to high recall rates in the US through the publication of updated national benchmarks,[2,5] the publicity of screening recommendation statements noting the harms of mammography screening,[20,21] and the expectation for lower recall as DBT disseminates.[22]

Notably, differences in unadjusted recall and cancer detection rates between radiologists who did vs. did not interpret DBT cannot be used to infer the effects of DBT on performance. We found differences in the characteristics of radiologists who did vs. did not interpret DBT, and differences in the characteristics of the women they screened, all of which may contribute to different unadjusted performance statistics. This is supported by our findings that radiologists who interpreted DBT had higher unadjusted recall and lower unadjusted cancer detection rates on DM before DBT interpretation compared to the cohort of radiologists who did not interpret DBT.

In our study DBT use was more frequent among among non-Hispanic white women and more commonly performend by fellowship-trained breast imaging specialists at academic medical centers. This suggests potential inequities in access to DBT, consistent with a recent study examining DBT adoption according to community-level socioeconomic resources,[23] and warrants further investigation.

We are aware of only one study examining DM performance after DBT experience, which reported increased recall and cancer detection rates for DM after vs. before DBT use in a single practice.[24] Notably, the six radiologists in the study had very low DM recall rates (average, 6.8%) and very low cancer detection rates (average 2.5 per 1000 exams) prior to DBT use. Our study, which included more than 100 radiologists from varied clinical settings across the US, suggests in contrast that DBT experience on average has little impact on DM recall rate or cancer detection rate.

Limitations

The large number of exams and radiologists from a geographically and racially diverse sample of academic and non-academic practice settings in the BCSC engenders greatly increased precision and representativeness compared to prior studies of DBT performance. However, the potential for selection bias must be considered. Our analysis of within-radiologist effects minimizes this source of bias by effectively using each radiologist as their own control. Collection of comprehensive exam-level data permitted the control of numerous factors known to be associated with recall and cancer detection rates, such as age, breast density, and time since last mammogram. Our inclusion of data from radiologists who began using DBT at a wide variety of points in time, as well as radiologists who did not interpret DBT, permitted control for secular trends.

A limitation of the study was our inability to examine radiologist-level cancer detection rates. However, we were able to evaluate the cancer detection rate separately for the subgroup of radiologists with a significant decrease in recall rate on DBT. While our results indicate that cancer detection overall is increased on DBT, future studies will be needed to examine variation in radiologist-level cancer detection rates. Additionally, given concerns about overdiagnosis in breast cancer[25,26] it is important to evaluate DBT-detected cancer characteristics. It remains unclear whether DBT is associated with a reduction in interval cancers, advanced stage cancers, or breast cancer mortality among women undergoing screening.[27–29]

Conclusions

In summary, we found that implementation of DBT leads to an overall reduction in radiologist recall rate and an overall increase in cancer detection rate. However, there is wide variability in the impact of DBT across radiologists. Thus, radiology practices should audit radiologist DBT performance statistics closely and consider additional DBT training for radiologists whose screening performance does not improve as expected. Policy makers and women should be informed that on average women undergoing DBT exams will experience a reduced recall rate and elevated cancer detection rate, though these benefits will vary substantially depending on the individual radiologist interpreting the exam.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Conflict of Interest Disclosures

Dr. Kerlikowske reported receiving grant support from Google Sciences outside the submitted work, and consulting with Grail on the STRIVE study. Dr. Lee has received grant funding from GE Healthcare for unrelated work and payment from the American College of Radiology for journal editorial board work. Dr. Miglioretti reported being a member of the Hologic Scientific Advisory Board. No other conflict of interest disclosures were reported.

Funding/Support

Research reported in this work was funded through a Patient-Centered Outcomes Research Institute (PCORI) award (PCS-1504-30370). Data collection for this research was additionally supported by the Breast Cancer Surveillance Consortium with funding from the National Cancer Institute (P01CA154292, U54CA163303), the Agency for Healthcare Research and Quality (R01 HS018366-01A1), and the University of Vermont Cancer Center with funds generously awarded by the Lake Champlain Cancer Research Organization (grant #032800). CI Lee is also supported in part by the American Cancer Society (126947-MRSG-1416001CPHPS). Cancer and vital status data collection was supported by several state public health departments and cancer registries (http://www.bcsc-research.org/work/acknowledgement.html.

Role of the Funder/Sponsor

The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions

We thank the participating women, mammography facilities, and radiologists for the data they have provided for this study. You can learn more about the BCSC at: http://www.bcsc-research.org/.

REFERENCES

1. Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L. Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med. 2016;164(4):256–267. [PubMed: 26756737]

2. Lehman CD, Arao RF, Sprague BL, et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. Radiology. 2017;283(1):49–58. [PubMed: 27918707]

3. Institute of Medicine. Improving Breast Imaging Quality Standards. Washington, DC: The National Academies Press;2005.

4. American College of Radiology. ACR BI-RADS® - Mammography 5th Edition ACR BI-RADS Atlas: Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology; 2013.

5. Carney PA, Sickles EA, Monsees BS, et al. Identifying minimally acceptable interpretive performance criteria for screening mammography. Radiology. 2010;255(2):354–361. [PubMed: 20413750]

6. Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. Radiology. 2006;241(1):55–66. [PubMed: 16990671]

7. Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. Medical physics. 2013;40(1):014301. [PubMed: 23298126]

8. Sechopoulos I. A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications. Medical physics. 2013;40(1):014302. [PubMed: 23298127]

9. U.S. Food and Drug Administration. MQSA National Statistics. https://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/FacilityScorecard/ucm595007.htm. Accessed April 11, 2019.

10. Marinovich ML, Hunter KE, Macaskill P, Houssami N. Breast Cancer Screening Using Tomosynthesis or Mammography: A Meta-analysis of Cancer Detection and Recall. J Natl Cancer Inst. 2018.

11. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol. 1997;169(4):1001–1008. [PubMed: 9308451]

12. Burnside ES, Lin Y, Munoz del Rio A, et al. Addressing the challenge of assessing physician-level screening performance: mammography as an example. PLoS One. 2014;9(2):e89418. [PubMed: 24586763]

13. Tice JA, Miglioretti DL, Li CS, Vachon CM, Gard CC, Kerlikowske K. Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer. J Clin Oncol. 2015;33(28):3137–3143. [PubMed: 26282663]

14. Miglioretti DL, Rutter CM, Geller BM, et al. Effect of breast augmentation on the accuracy of mammography and cancer characteristics. JAMA. 2004;291(4):442–450. [PubMed: 14747501]

15. Houssami N, Abraham LA, Miglioretti DL, et al. Accuracy and outcomes of screening mammography in women with a personal history of early-stage breast cancer. JAMA. 2011;305(8):790–799. [PubMed: 21343578]

16. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? Stat Med. 2009;28(2):221–239. [PubMed: 19012297]

17. Gilbert FJ, Tucker L, Young KC. Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. Clin Radiol. 2016;71(2):141–150. [PubMed: 26707815]

18. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA. 2014;311(24):2499–2507. [PubMed: 25058084]

19. Melnikow J, Fenton JJ, Miglioretti D, Whitlock EP, Weyrich MS. Screening for Breast Cancer with Digital Breast Tomosynthesis, AHRQ Publication No. 14–05201-EF-2. Rockville, MD: Agency for Healthcare Research and Quality;2016.

20. Siu AL, Force USPST. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2016;164(4):279–296. [PubMed: 26757170]

21. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA. 2015;314(15):1599–1614. [PubMed: 26501536]

22. Houssami N, Miglioretti DL. Digital Breast Tomosynthesis: A Brave New World of Mammography Screening. JAMA oncology. 2016.

23. Richman IB, Hoag JR, Xu X, et al. Adoption of Digital Breast Tomosynthesis in Clinical Practice. JAMA Intern Med. 2019.
24. DiPrete O, Lourenco AP, Baird GL, Mainiero MB. Screening Digital Mammography Recall Rate: Does It Change with Digital Breast Tomosynthesis Experience? Radiology. 2017:170517.
25. Narod S. Breast cancer: The importance of overdiagnosis in breast-cancer screening. Nat Rev Clin Oncol. 2016;13(1):5–6. [PubMed: 26573419]
26. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. N Engl J Med. 2016;375(15):1438–1447. [PubMed: 27732805]
27. Hovda T, Holen AS, Lang K, et al. Interval and Consecutive Round Breast Cancer after Digital Breast Tomosynthesis and Synthetic 2D Mammography versus Standard 2D Digital Mammography in BreastScreen Norway. Radiology. 2019:191337.
28. McDonald ES, Oustimov A, Weinstein SP, Synnestvedt MB, Schnall M, Conant EF. Effectiveness of Digital Breast Tomosynthesis Compared With Digital Mammography: Outcomes Analysis From 3 Years of Breast Cancer Screening. JAMA oncology. 2016.
29. Bahl M, Gaffney S, McCarthy AM, Lowry KP, Dang PA, Lehman CD. Breast Cancer Characteristics Associated with 2D Digital Mammography versus Digital Breast Tomosynthesis for Screening-detected and Interval Cancers. Radiology. 2018;287(1):49–57. [PubMed: 29272213]

KEY POINTS

Question:

Is digital breast tomosynthesis (DBT) associated with improved radiologist-level breast cancer screening performance?

Findings:

In this Breast Cancer Surveillance Consortium cohort, DBT was associated with an overall 15% decrease in recall rate and a 21% increase in cancer detection rate compared to digital mammography (DM). Recall rates were significantly lower on DBT compared to DM exams interpreted prior to DBT use for 35.7% of radiologists and significantly higher for 14.3%; 50.0% had no statistically significant change in recall rate.

Meaning:

DBT is associated with a clinically important overall reduction in recall rate and an overall increase in cancer detection rate; however, there is wide variability with many radiologists showing no significant improvement in recall rate with DBT.

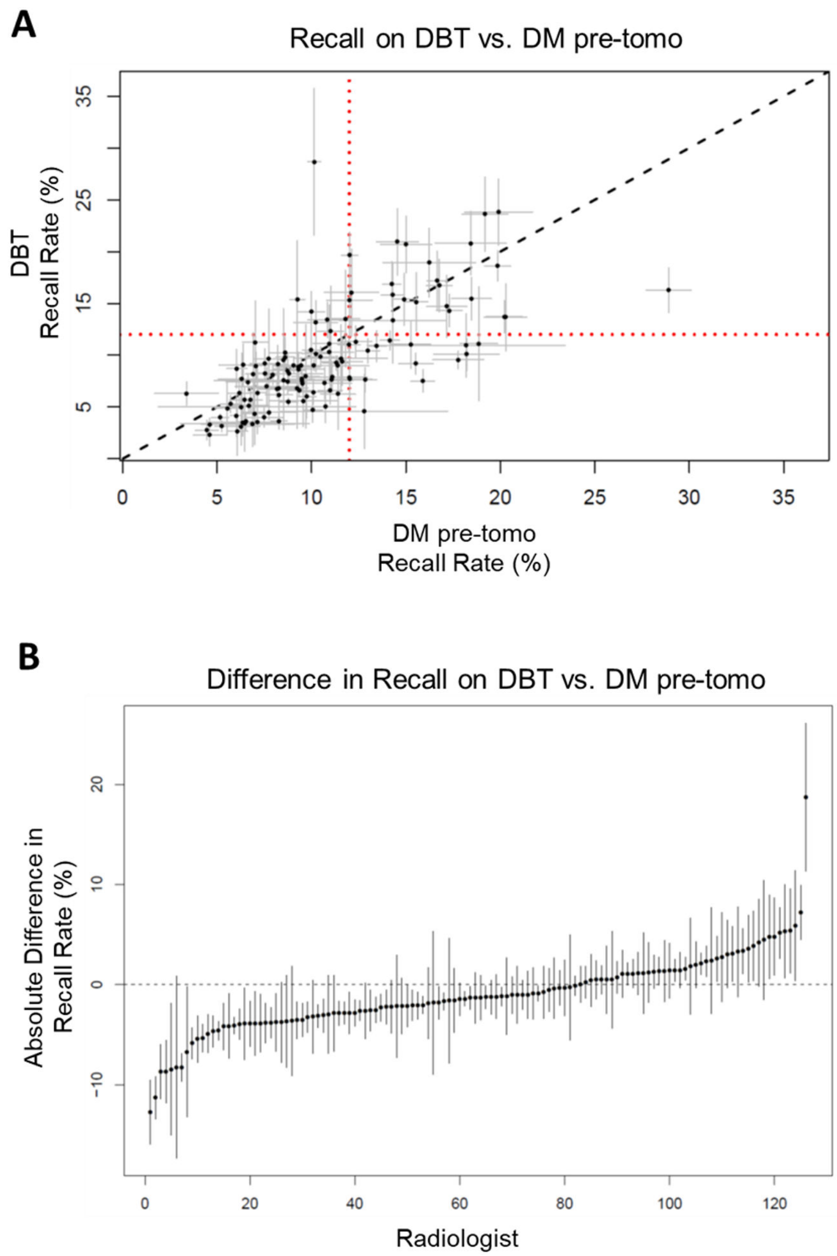

Figure 1.

Screening recall rates on mammography exams with digital breast tomosynthesis (DBT) compared to digital mammography exams (DM) prior to the use of digital breast tomosynthesis (“DM pre-DBT”). (A) Scatterplot of raw recall rate on DBT exams vs. DM pre-DBT. Each point represents a single radiologist; error bars depict 95% confidence intervals; the dashed black line represents the points at which the recall rates are equal in the comparison groups; the dashed red lines represents the expert-recommended upper threshold for recall rate of 12%. (B) Distribution of the multivariable-adjusted difference in recall rate on DBT vs. DM pre-DBT, by radiologist. Error bars represent the width of the 95% confidence interval. The horizontal dashed line represents no difference in recall rate between comparison groups.

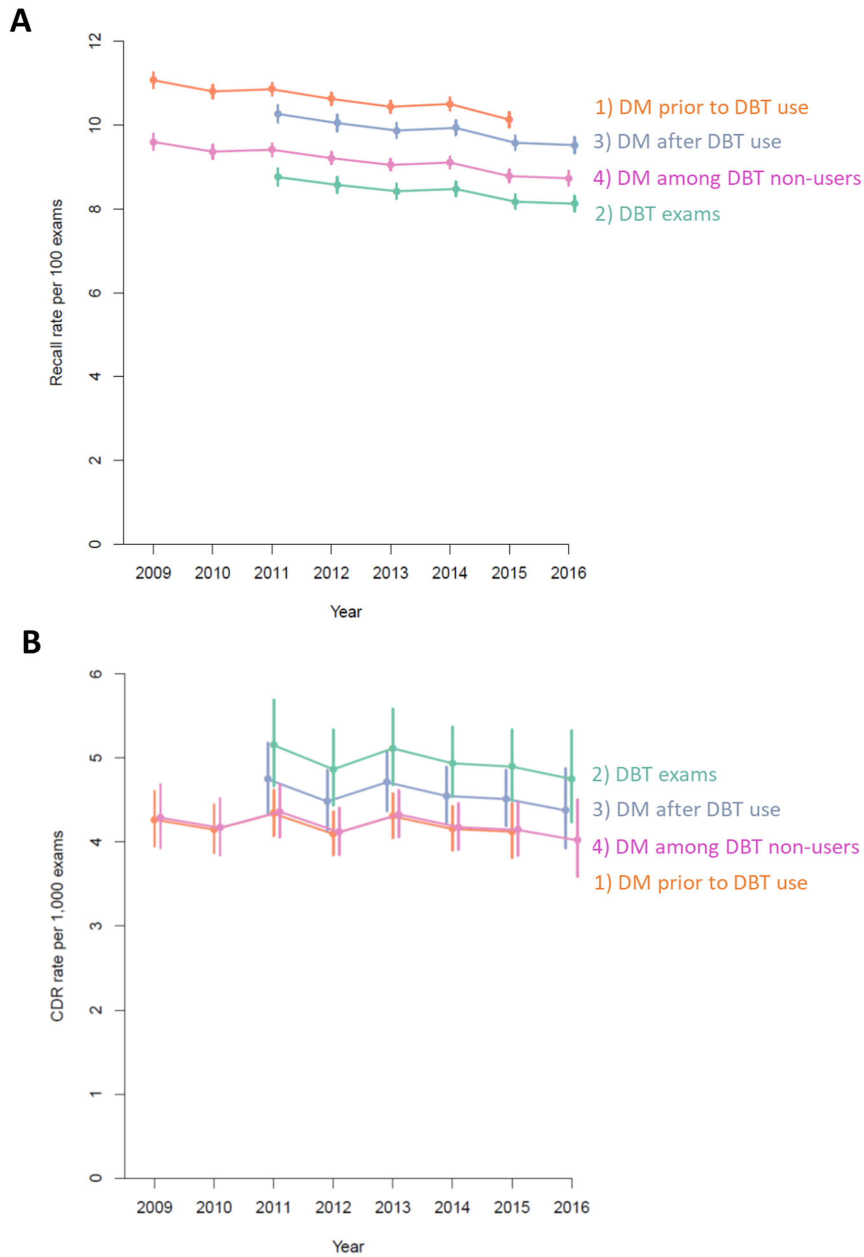

Figure 2.

Multivariable-adjusted recall rate (A) and cancer detection rate (B) by calendar year according to comparison group, adjusted for the average exam characteristics profile and radiologist-level effects. Error bars depict 95% confidence intervals. Partial year data for 2017 is not shown. DBT, digital breast tomosynthesis; DM, digital mammography.

Table 1.

Characteristics of 2,252,065 screening exams interpreted during 2009–2017 by 198 radiologists from 104 facilities in the Breast Cancer Surveillance Consortium. Percentages are column percentages among non-missing; percent of radiologists with missing data is indicated in parentheses.

	Radiologists Who Interpreted DBT									
	All exams		**DM prior to DBT use**		**DBT exams**		**DM after DBT use**		**DM Exams Among Radiologists Who Did Not Interpret DBT**	
	N=198 radiologists N=2,252,065 exams		**N=126 radiologists N=945,635 exams**		**N=126 radiologists N=251,384 exams**		**N=114 radiologists N=330,730 exams**		**N=72 radiologists N=724,316 exams**	
	No.	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Age, years										
<40	33,141	1%	17,880	2%	4,033	2%	3,219	1%	8,009	1%
40–49	543,631	24%	251,115	27%	66,389	26%	71,825	22%	154,302	21%
50–59	710,934	32%	294,155	31%	80,978	32%	102,861	31%	232,940	32%
60–69	601,484	27%	228,322	24%	64,278	26%	91,254	28%	217,630	30%
70–79	283,747	13%	116,205	12%	29,042	12%	46,916	14%	91,584	13%
80+	79,128	4%	37,958	4%	6,664	3%	14,655	4%	19,851	3%
Race, Ethnicity										
White, non-Hispanic	1,448,981	67%	569,594	63%	190,399	80%	200,739	65%	488,249	70%
Black, non-Hispanic	252,790	12%	146,761	16%	17,361	7%	44,887	14%	43,781	6%
Hispanic	133,049	6%	66,742	7%	11,287	5%	19,339	6%	35,681	5%
Asian/Pacific Islander	268,871	12%	102,478	11%	13,872	6%	39,152	13%	113,369	16%
Mixed/other	49,697	2%	18,480	2%	5,666	2%	6,521	2%	19,030	3%
Missing	98,677	(4%)	41,580	(4%)	12,799	(5%)	20,092	(6%)	24,206	(3%)
First degree family history										
Yes	348,572	16%	140,347	15%	37,537	16%	43,440	14%	127,248	18%
No	1,833,239	84%	779,878	85%	200,679	84%	275,057	86%	577,625	82%
Missing	70,254	(3%)	25,410	(3%)	13,168	(5%)	12,233	(4%)	19,443	(3%)
History of breast biopsy										
No	1,768,237	81%	747,166	81%	187,195	79%	248,159	78%	585,717	83%
Yes	364,882	17%	149,152	16%	44,926	19%	48,906	15%	121,898	17%
Missing	118,946	(5%)	49,317	(5%)	19,263	(8%)	33,665	(10%)	16,701	(2%)
Breast density										
Almost entirely fatty	225,627	11%	98,885	11%	22,491	9%	40,500	13%	63,751	10%

	Radiologists Who Interpreted DBT									
	All exams		DM prior to DBT use		DBT exams		DM after DBT use		DM Exams Among Radiologists Who Did Not Interpret DBT	
	N=198 radiologists N=2,252,065 exams		N=126 radiologists N=945,635 exams		N=126 radiologists N=251,384 exams		N=114 radiologists N=330,730 exams		N=72 radiologists N=724,316 exams	
	No.	%	No.	%	No.	%	No.	%	No.	%
Scattered fibroglandular densities	888,904	42%	372,349	42%	107,162	44%	151,315	47%	258,078	40%
Heterogeneously dense	811,501	39%	338,317	38%	91,900	38%	107,562	33%	273,722	42%
Extremely dense	173,226	8%	79,282	9%	20,167	8%	23,570	7%	50,207	8%
Missing	152,807	(7%)	56,802	(6%)	9,664	(3%)	7,783	(2%)	78,558	(11%)
BCSC 5 year risk[a]										
< 1.00%	618,484	32%	285,752	35%	69,465	31%	96,842	33%	166,425	28%
1.00 – 1.66%	769,204	40%	318,713	40%	89,809	40%	120,275	42%	240,407	40%
1.67 – 2.49%	375,685	20%	144,716	18%	45,516	20%	52,902	18%	132,551	22%
2.50 – 3.99%	138,100	7%	49,575	6%	16,796	7%	17,502	6%	54,227	9%
≥4.00%	19,718	1%	6,485	1%	2,650	1%	2,286	1%	8,297	1%
Missing	330,874	(15%)	140,394	(15%)	27,148	(11%)	40,923	(12%)	122,409	(17%)
Time since last mammogram										
1–2 years	1,840,232	84%	768,518	83%	209,197	85%	270,286	83%	592,231	85%
3–4 years	128,920	6%	50,427	5%	11,500	5%	17,489	5%	49,504	7%
5+ years	74,880	3%	31,244	3%	7,399	3%	10,709	3%	25,528	4%
No previous mammogram	145,148	7%	71,602	8%	16,663	7%	26,947	8%	29,936	4%
Missing	62,885	(3%)	23,844	(2%)	6,625	(3%)	5,299	(2%)	27,117	(4%)

DM, digital mammography; DBT, digital breast tomosynthesis.

[a]Estimated risk based on the Breast Cancer Surveillance Consortium 5-year risk model (version 2).[13]

Table 2.

Characteristics of 198 radiologists included in the study from 104 facilities in the Breast Cancer Surveillance Consortium. Percentages are column percentages among non-missing; percent of radiologists with missing data is indicated in parentheses.

	Radiologists who did not interpret DBT (N=72)		Radiologists who interpreted DBT: All (N=126)		Statistically significant decrease in recall rate on DBT (N=45)		No change in recall rate on DBT (N=63)		Statistically significant increase in recall rate on DBT (N=18)		Chi-square P-value*
	No.	%	No.	%	No.	%	No.	%	No.	%	
Breast imaging specialist											0.57
No	51	86%	76	73%	26	67%	36	77%	14	78%	
Yes	8	14%	28	27%	13	33%	11	23%	4	22%	
Missing	13	(18%)	22	(18%)	6	(13%)	16	(25%)	0	(0%)	
Fellowship trained in breast imaging											0.73
No	54	92%	76	72%	29	73%	33	69%	14	78%	
Yes	5	9%	30	28%	11	28%	15	31%	4	22%	
Missing	13	(18%)	20	(16%)	5	(11%)	15	(24%)	0	(0%)	
Practice at academic medical center											0.59
No	71	99%	100	80%	36	80%	48	77%	16	89%	
Yes	1	1%	25	20%	9	20%	14	23%	2	11%	
Missing	0	(0%)	1	(1%)	0	(0%)	1	(2%)	0	(0%)	
Practice type											0.65
Multi-specialty breast center	42	58%	63	50%	20	44%	33	53%	10	56%	
Radiology practice	30	42%	62	50%	25	56%	29	47%	8	44%	
Missing	0	(0%)	1	(1%)	0	(0%)	1	(2%)	0	(0%)	
Facility Location											0.37
Hospital	40	56%	113	90%	42	93%	56	89%	15	83%	
Office not in hospital	3	4%	11	9%	2	4%	6	10%	3	17%	
Mobile van	16	22%	0	0%	0	0%	0	0%	0	0%	
Other	13	18%	1	1%	1	2%	1	2%	0	0%	
Missing	0	(0%)	1	(1%)	0	(0%)	0	(0%)	0	(0%)	
Profit status of practice											0.96
For profit	8	19%	41	38%	17	40%	19	37%	5	36%	

	Radiologists who did not interpret DBT (N=72)		All (N=126)		Radiologists who interpreted DBT: Statistically significant decrease in recall rate on DBT (N=45)		Radiologists who interpreted DBT: No change in recall rate on DBT (N=63)		Radiologists who interpreted DBT: Statistically significant increase in recall rate on DBT (N=18)		Chi-square P-value*
	No.	%	No.	%	No.	%	No.	%	No.	%	
Non-profit	35	81%	68	62%	26	61%	33	64%	9	64%	
Missing	29	(40%)	17	(14%)	2	(4%)	11	(18%)	4	(22%)	
Annual screening volume											0.05
100–999	13	18%	32	25%	7	16%	23	37%	2	11%	
1,000 – 1,999	26	36%	48	38%	17	38%	24	38%	7	39%	
2,000 – 4,999	29	40%	43	34%	20	44%	14	22%	9	50%	
5,000+	4	6%	3	2%	1	2%	2	3%	0	0%	
Percent of screens that were DBT (vs. DM) after initiation of DBT											0.001
0–24.9%	NA	NA	32	25.4	4	8.9	22	35%	6	33%	
25–49.9%	NA	NA	40	31.7	14	31.1	21	33%	5	28%	
50–74.9%	NA	NA	27	21.4	17	37.8	6	10%	4	22%	
75–99.9%	NA	NA	15	11.9	4	8.9	9	14%	2	11%	
100%	NA	NA	12	9.5	6	13.3	5	8%	1	6%	

DBT, digital breast tomosynthesis; DM, digital mammography; NA, not applicable.

* Chi-square test comparing distribution of characteristics across radiologist groups that had a statistically significant decrease vs. no change vs. statistically significant recall rate on DBT compared to pre-DBT DM screens.

Table 3.

Recall rate and cancer detection rate by screening mammography modality among 126 radiologists in the Breast Cancer Surveillance Consortium who began interpreting DBT during the study period (2009–2017).

	DM exams interpreted prior to DBT use	DBT exams	DM exams interpreted after DBT use
Recall rate			
N, exams	945,635	251,384	330,730
N, radiologists	126	126	114
Median unadjusted recall rate, % (25^{th}-75^{th} percentile)	10.0 (7.5–13.0)	8.8 (6.3–11.3)	9.2 (7.0–11.9)
Fraction with unadjusted recall ≤12%	65.9%	76.2%	77.2%
Adjusted RR for recall (95% CI) *	1.00 (ref)	0.85 (0.83–0.87)	1.01 (0.99–1.03)
Cancer detection rate †			
N, exams	930,509	186,863	284,137
N, radiologists	126	124	111
Mean unadjusted cancer detection rate , per 1000 exams (95% CI)	4.7 (4.6–4.8)	5.3 (5.0–5.7)	5.1 (4.8–5.3)
Adjusted RR for cancer detection (95% CI) *	1.00 (ref)	1.21 (1.11–1.33)	1.06 (0.98–1.14)

CI, confidence interval; DM, digital mammography; DBT, digital breast tomosynthesis; RR, relative risk.

* Multivariable-adjusted model is adjusted for secular trends (exam year) and woman-level factors including age, race/ethnicity, first degree family history of breast cancer, history of breast biopsy, BI-RADS breast density, BCSC 5-year risk, time since last mammogram, history of prior DBT imaging, and BCSC registry site.

† Statistics for cancer detection rate are limited to exams with 365 days of follow-up for cancer diagnoses.