Title

Feasibility and Acceptability of Conducting a Randomized Clinical Trial Designed to Improve Interpretation of Screening Mammography

Journal

Academic Radiology, 20(11)

ISSN

1076-6332

Authors

Carney, Patricia A
Bogart, Andy
Sickles, Edward A
et al.

Publication Date

2013-11-01

DOI

10.1016/j.acra.2013.08.017

Peer reviewed

Published in final edited form as:
Acad Radiol. 2013 November ; 20(11): 1389–1398. doi:10.1016/j.acra.2013.08.017.

Feasibility and Acceptability of Conducting a Randomized Clinical Trial Designed to Improve Interpretation of Screening Mammography

Patricia A. Carney, PhD[1], **Andy Bogart, MS**[2], **Edward A. Sickles, MD**[3], **Robert Smith, PhD**[4], **Diana SM Buist, PhD**[2], **Karla Kerlikowske, MD**[5], **Tracy Onega, PhD**[6], **Diana L. Miglioretti, PhD**[7,2], **Robert Rosenberg, MD**[8], **Bonnie C. Yankaskas, PhD**[9], and **Berta M. Geller, EdD**[10]

[1] Departments of Family Medicine and Public Health and Preventive Medicine, Oregon Health & Science University, Portland, OR

[2] Group Health Research Institute, Group Health Cooperative, Seattle, WA

[3] Department of Radiology, University of California, San Francisco, San Francisco, CA

[4] Cancer Control Science Department, American Cancer Society, Atlanta, GA

[5] Departments of Medicine and Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA

[6] Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH

[7] Department of Public Health Sciences, University of California Davis School of Medicine, Davis, WA

[8] Department of Radiology, University of New Mexico, Albuquerque, NM

[9] Department of Radiology, University of North Carolina, Chapel Hill, NC

[10] Department of Family Medicine, University of Vermont, Burlington, VT

Abstract

Purpose—To describe recruitment, enrollment and participation in a study of U.S. radiologists invited to participate in a randomized controlled trial of two continuing medical education interventions designed to improve interpretation of screening mammography.

Methods—We collected recruitment, consent, and intervention-completion information as part of a large study involving radiologists in California, Oregon, Washington, New Mexico, New Hampshire, North Carolina, and Vermont. Consenting radiologists were randomized to receive either a one-day live, expert-led educational session; a self-paced DVD with similar content; or to a control group (delayed intervention). The impact of the interventions was assessed using a pre- and post-intervention test set design. All activities were IRB-approved and HIPAA compliant.

Correspondence to: Patricia Carney, PhD Department of Family Medicine Oregon Health & Science University 3181 SW Sam Jackson Park Rd. Mail Code: FM Portland, OR 97239 Phone: 503-494-9049 Fax: 503-494-2746 carneyp@ohsu.edu.

Results—Of 403 eligible radiologists, 151/403 (37.5%) consented to participate in the trial and 119/151 (78.8%) completed the pre–intervention test set, leaving 119 available for randomization to one of the two intervention groups or to controls. Female radiologists were more likely than males to consent to and complete the study (p= 0.03). Consenting radiologists who completed all study activities were more likely to have been interpreting mammography for ≤10 years compared to radiologists who consented and did not complete all study activities or did not consent at all. The live intervention group was more likely to report their intent to change their clinical practice as a result of the intervention compared to those who received the DVD (50% versus 17.6%, p=0.02). The majority of participants in both interventions groups felt the interventions were a useful way to receive CME mammography credits

Conclusions—Community radiologists found interactive interventions designed to improve interpretative mammography performance acceptable and useful for clinical practice. This suggests CME credits for radiologists should, in part, be for examining practice skills.

Introduction

Continuing medical education (CME) has traditionally been a requirement for maintaining qualifications for practicing physicians (1). Physicians who interpret mammography are required by the Mammography Quality Standards Act (MQSA) to obtain at least 15 hours of Category 1 CME units in mammography every 36 months to maintain their qualifications (2). Justification for continuing CME activities under MQSA is based on a belief that gains in knowledge will lead to improved patient care and outcomes (3). However, despite the significant level of participation and resources applied to CME, there are two persistent concerns. First, conventional, lecture-based CME may have little if any effect on physician performance (3-6). Second, 20 years after Congress passed MQSA, there still is a sizable gap between actual and ideal interpretative performance (7, 8).

In 1992, the definition of traditional CME had expanded beyond classic passive lectures or grand rounds, as physicians and CME providers were undertaking more complex learning activities such as computer-based simulations using actual patient problems, reading materials, and visits to practice sites from health care professionals trained to improve performance using academic detailing (3). Several such approaches have been described as positive interventions because they prepared physicians for further learning and improvements in clinical practice (3). In addition, subsequent studies (5, 6, 9, 10) of more discrete interventions consistently identified three important features of effective CME: (1) assessment of learning needs is a necessary precursor to effective CME; (2) the importance of interaction among physician-learners with opportunities to practice the skills learned; and (3) the importance of multifaceted educational activities (5, 6, 9-12).

Several studies have tested approaches to improve interpretive performance of screening mammography, the most of which combined several strategies, including performance data review, participation in a self-assessment and case review program, and increasing interpretive volume (13-16). What is less well understood in educational intervention research is how feasible it is to engage clinical practitioners to participate in complex educational research. Understanding the characteristics of those who consent to educational

research as well as the characteristics of those who complete all study components compared with those who drop out can assist in tailoring future recruitment efforts, and in interpreting findings from educational interventions.

We conducted an interpretive skills assessment using mammography test sets before and after testing two educational strategies designed to improve interpretive performance of screening mammography relative to a control group. In this paper, we report what we learned about the feasibility and acceptability of conducting a large complex randomized controlled trial to assess educational interventions.

Methods

This study enrolled radiologists to: 1) complete a brief survey and complete one of four mammography pre-intervention test sets designed to assess their baseline performance, 2) be randomized to receive one of two interventions or serve as a control group (delayed intervention), 3) complete the intervention if randomized to one, and 4) complete a post-intervention test set. The larger study is described in detail elsewhere (17, 18). Briefly, we developed four image-based test sets designed to assess interpretative performance at baseline, conducted a randomized controlled trial to evaluate two educational interventions designed to improve interpretation of screening mammography, and designed a single test-set to test performance post-intervention. A third study arm served as a control group. The development of the test sets is described elsewhere (17), and the content of the interventions was based on what we learned about participants' performance on the pre-intervention test sets, which were administered the year before the interventions were developed and deployed. The interventions included a self-paced DVD and a live expert-led 8 hour educational session that included review of 40 cases (18-20). If radiologists had a compelling reason for being unable to attend the live intervention after the initial randomization, they were re-randomized to either the DVD group or the Control Group. This occurred for 13 participants (6 moved from Live to DVD, and 7 moved from Live to Control, see **Figure 2**). One other participant was mistakenly invited to attend the live intervention, despite having been randomized to the DVD group. This person was reassigned to the live intervention. To evaluate the interventions, we compared participants' performance on a post-intervention test set administered at least 90 days after the interventions were completed. The impact of the interventions on radiologist performance is reported elsewhere (18).

Study population

During enrollment, which occurred in 2009 and 2010, we invited 403 radiologists to participate. Eligibility included those who interpreted mammograms at a facility contributing to a National Cancer Institute Breast Cancer Surveillance Consortium (BCSC) mammography registry (21) between January 2005 and December 2006. We also invited 103 non-BCSC radiologists from Oregon; Puget Sound, WA, North Carolina, San Francisco, and New Mexico. As an incentive to participate, all participants received up to 24 Category 1 CME credits through the University of Vermont for completing the three components: 1) the pre-intervention test set, 2) either of live or DVD intervention being tested (or delayed

DVD intervention for the control group), and the post-intervention test set. Potential participants were notified that they could receive up to 24 AMA PRA Category 1 credits Continuing Medical Education (CME).

Each BCSC registry and the Statistical Coordinating Center (SCC), where analyses were performed, received IRB approval for all study activities, including active consent to enroll radiologists and perform analytic studies. All registries and the SCC follow procedures that are Health Insurance Portability and Accountability Act (HIPPA) compliant to obtain films and patient information and also have received a Federal Certificate of Confidentiality and other protections for the identities of women, physicians, and facilities that are related to the films used in this research (22).

Data collection

Study coordinators at each site were provided with a tracking database, which was used to maintain records of all study activities including participant recruitment, administration of the pre- and post-intervention test sets, and all activities related to randomization and implementation of study interventions. Data for this study were collected from several sources and availability varied by group (**Figure 1**). Characteristics of eligible radiologists were obtained using survey data from a previous study (23). Among consenting radiologists who completed all aspects of the study, we were able to obtain demographic and practice characteristics on 81/102 (79.4%). Among those who consented but did not complete the study activities, we were able to obtain similar data on 31/49 (63.3%). Among the radiologists who did not consent to take part, we were able to obtain data on 95/252 (37.3%). The majority of radiologists, those in the BCSC, in the study had long term relationships with the investigators in this study from their contributions of interpretation data to the respective registries, in some cases over several years.

Interpretive performance data from the respective BCSC mammography registries (18) was available for 82 of 102 (80.4%) consenting radiologists who completed all study activities, 42/49 (85.7%) consenting radiologists who did not complete all study activities, and 166 of 252 (65.9%) of radiologists who chose not to take part in the study. The BCSC data includes linkages to breast pathology labs and/or state cancer registries that allow for accurate calculations of standard performance measures (21). Questions related to both interpreting the test sets and satisfaction with the assigned interventions were completed by the majority of participants (50/70 of those randomized to the Live or DVD interventions, and 35/49 of those randomized to control) as part of the CME, which allowed us to award CME credit for study activities. Column headers in each table presented here provide the number of radiologists for whom data were available, (shown as a numerator), as well as the number of radiologists who comprised the category overall (shown as a denominator). Percentages reported in each table are calculated out of the number of radiologists with available data. Performance measures from the BCSC data were calculated using standard definitions (24).

Data Analysis

We constructed descriptive tables to compare measures from each data source across three groups; 1) consenting radiologists who completed all aspects of the study; 2) consenting

radiologists who did not complete all study hypotheses; and 3) radiologists who were invited but did not consent to participate in the study. Tables are specific to data sources and present data only from the subset of radiologists for whom valid data were available. As some data were sparse for some data sources, we uniformly compared categorical items across the three groups using Fisher's exact test. Continuous items were compared using the Kruskall-Wallis test, an extension of the nonparametric Wilcoxon Rank Sum test for comparing more than two groups (25). All data management and analysis was conducted using SAS software, version 9.2.

Results

Of the 403 eligible radiologists we identified, 151/403 (37.5%) consented to the trial and 119/151 (78.8%) completed the pre-intervention test set (**Figure 2**), leaving 119 available for randomization to an intervention group. Of the 26 physicians who represented the final group assigned to the expert-led live intervention, all attended (26/26,100%), and of the 44 randomized to the self-paced DVD, 41 completed it (93.2%). Twenty-five of twenty-six participants (96.2%) who attended the live intervention completed the post-intervention test: 37/41 (90.2%) who completed the DVD intervention completed the post intervention test set and 40/49 (81.6%) of the control physicians completed it.

We found that female radiologists were more likely than males to consent for, and complete the study (**Table1**). Consenting radiologists who completed study activities were more likely to have been interpreting mammography for 10 years or less compared to radiologists who consented and did not complete the study or did not consent at all, the latter group tending to have been in practice for 20 or more years (**Table 1**). We found no significant differences in the three groups based on academic affiliation, fellowship training, percent time spent in breast imaging or CME preferences or attitudes.

We found no significant differences between clinical interpretive performance, including volume, number of screening mammograms associated with a cancer diagnosis in the five years prior to the intervention, sensitivity, specificity, recall rate, PPV 1, and cancer detection rate among the radiologists who consented and completed all study activities, those who consented but did not complete study activities and those who did not consent (**Table 2**). Median sensitivity was above 86% for all categories of participating and non-participating radiologists (**Table 2)**. Median specificity varied between groups from 90.7 to 91.3, median recall rate varied from 9.0 to 9.7%, median PPV1 varied from 4.0% to 4.7% and median cancer detection rate per 1,000 exams varied from 3.8 to 4.4, with no statistically significant differences according to study group.

Overall, a majority of participants (56.3%) reported that the types of abnormal findings used in the four pre-intervention test sets were definitely representative of those they see in clinical practice (**Table 3**), regardless of the prevalence of cancers in their assigned test set. The vast majority (>85%) thought their assigned test set was somewhat or definitely useful for evaluating their skills. At least 70% thought they would change their clinical practice because of what they learned from interpreting the pre-intervention test sets. More than 94% found the feedback they received as part of the pre-intervention test set exercise helpful to

improving their practice, and over 97% thought this was a useful way to receive CME mammography credits. We found no statistically significant differences in responses to these satisfaction questions according to study group assignment.

Participant satisfaction with their assigned interventions is shown in **Table 4**. Nearly 77% of those in the live intervention reported the instructional methods were definitely appropriate for their learning style compared to 32.4% of those assigned to the DVD (p=0.002). Those assigned to the live intervention were also more likely to report their intent to change their clinical practice as a result of the intervention received compared to those who received the DVD (50% versus 17.6%, p=0.02). The majority of participants in both study groups felt their assigned interventions were a useful way to receive CME mammography credits (**Table 4**).

Table 5 presents information on satisfaction with the follow-up test set by study group assignment. Responses to the questionnaire regarding interpreting the test set were similar to those of the pre-intervention test set except for one indicator. Radiologists in the control group were more likely to report that getting feedback on questions answered as part of the interpreting the test set and how their performance compared with peers was definitely helpful compared to those in the other intervention groups (74.2% versus 34.8% and 35.3%, P=0.005).

Discussion

This study is one of only two studies we are aware of that have used randomized controlled trial designs to test the impact of interventions planned to improve the interpretive performance of screening mammography (26). Other studies were based on a pre-post evaluation design rather than a randomized controlled design, and used review of performance data, participation in a self-assessment and case review program, and increasing interpretive volume (10-13, 27) to assess changes in clinical practice. Pre-post designs are weaker than an RCT because it is not possible to fully attribute findings to an intervention without the benefit of a control or comparison group taking part in the same evaluation activities as the intervention group.

We succeeded in attracting nearly 38% of eligible radiologists to the study during the recruitment/enrollment phase. More than 44% of eligible radiologists did not respond to any of the invitations we extended for study participation, and 18% actively refused to take part. We are unable to determine whether this response was due to lack of time or interest or whether there was not a need for the CME credits offered as part of the study, or both. We asked for an estimated 24 hours of their time for this study, which is a significant time commitment. If lack of time was the reason for not participating, it may be that these physicians have readily available opportunities to fulfill CME requirements and felt it unnecessary to undertake the professional opportunities offered in this study. The bias in this hypothetical example is that physicians who do not need CME credits would be less likely to be represented in community-based clinical research testing the effectiveness of interventions to improve clinical research and therefore reduce the generalizability of findings. It also may be that these radiologists perceive there is no need to improve their

interpretive performance, as they believe they are performing well already. In fact, there is some evidence that those who did not participate had higher, although not statistically higher, clinical sensitivity than those who completed the study. Alternatively, some health professionals may prefer to avoid having their performance measured, for fear that it may not live up to their own and others perception of their competence. Any of these reasons are concerning because the most robust effectiveness research captures a representative sample of all eligible potential participants.

In our study, we evaluated the characteristics of radiologists who consented and completed the study, those who started and did not complete the study and those who did not consent, so we could understand the generalizability of our findings. The good news is that once enrolled, nearly 79% completed the pre-intervention test set and went on to randomization to one of two interventions or a control group. Once randomized, over 90% participated in their assigned intervention groups and completed the post-intervention test set. Nearly 82% of those who agreed to participate and assigned to the control group completed the post-intervention test set, which is a notable achievement and demonstrates that this type of research is feasible. As an incentive, we offered the self-paced DVD to the physicians in the comparison group *after* they completed the post-intervention test set so we could provide the full 24 hours of CME credit to them as we did for those assigned to the intervention groups, which likely increased our ability to capture follow-up data. These findings suggest that this intervention study was both feasible and acceptable to participants. Of note is that it took a number of contacts to encourage radiologists to complete study activities. Though not specifically reported in results because we could not obtain accurate counts, these included mail and telephone follow-up and in some cases, sending a small gift, such as chocolates or coffee cards to encourage completion of study activities.

Our interpretation of the finding that participants in the control group valued the peer comparison feedback they received as part of the post intervention test set was higher because this was the only feedback they received as part of the study until they were provided the DVD after they completed interpreting the post intervention test set. Those in the other two intervention groups may have rated this variable lower because they found greater value from the feedback they received as part of the educational interventions themselves. Of note is that radiologists appear to greatly value feedback about their performance.

One of the greatest challenges in all intervention studies is capturing information on individuals who decline to participate. We were fortunate to have information on the characteristics of nearly 40% of non-responders through their participation in a prior survey and on the interpretive performance of 65% of non-responders through their participation in the BCSC (21). These data indicated that the only significant differences between non-participants and participants were gender and number of years interpreting mammography. Female physicians were more likely to enroll and complete study activities compared to male radiologists, a finding that is consistent with another randomized controlled study this team conducted (26). We speculate that this might have occurred because female physicians are almost three times more likely to be in part time practice than their male counterparts (28) and thus have more time to undertake this activity. Additionally, we found that

radiologists who were newer to clinical practice were more likely to complete all components than those who did not enroll or enrolled and did not complete the intervention. It may be that these more junior physicians might either need more CME credits or be more open to educational activities than those who have been in practice longer, or that part time physicians have less access to CME funds to pay for their required CME activities. It may be that physicians who have been in practice longer than 20 years are closer to retirement and thus less likely to be interested in enrolling in such a study. In any case, these factors may limit the generalizability of our intervention results.

A strength of this study is that we were able to collect detailed information on mainly community-based radiologists who participated in a complex, time intensive randomized controlled trial. In addition, we were able to characterize the traits of physicians who enrolled and completed all study activities compared to those who did not complete them. Next steps in this line of research include disseminating the intervention, which showed clinically useful effects (15), and understanding the uptake of the intervention in a dissemination study.

In conclusion, most community-radiologists who enroll in a randomized controlled trial designed to test the effectiveness of interventions to improve clinical practice complete all study activities: however, the challenge is getting them to enroll. Evaluating the characteristics of those who choose not to participate is important to fully understanding the generalizability of intervention results.

Acknowledgments

This work was supported the American Cancer Society and made possible by a generous donation from the Longaberger Company's Horizon of Hope Campaign (SIRGS-07-271-01, SIRGS-07-272-01, SIRGS-07-274-01, SIRGS-07-275-01, SIRGS-06-281- 01, ACS A1-07-362). It was also supported by the National Cancer Institute (NCI) Breast Cancer Surveillance Consortium (BCSC) (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040, HHSN261201100031C) and another NCI funded study (1R01 CA107623). The collection of cancer data used in this study was supported in part by several state public health departments and cancer registries throughout the U.S. For a full description of these sources, please see: http://breastscreening.cancer.gov/work/acknowledgement.html.The authors had full responsibility in the design of the study, the collection of the data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, and the writing of the manuscript. We thank the participating women, mammography facilities, and radiologists for the data they have provided for this study. A list of the BCSC investigators and procedures for requesting BCSC data for research purposes are provided at: http://breastscreening.cancer.gov/.

References

1. Davis D, O'Brien MT, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A. Impact of Formal Continuing Medical Education: Do Conferences, Workshops, Rounds, and Other Traditional Continuing Education Activities Change Physician Behavior or Health Care Outcomes? JAMA. 1999; 282(9):867–874. [PubMed: 10478694]
2. Federal REgister. Vol. 58. Government Printing Office1992; Washington, DC: Dec 21. 1993 Mammography quality standards act of 1992. Mammography facilities requirement for accrediting bodies, and quality standards and certifying requirements: interim rules (21 CFR 900).; p. 57558-72.1992. Report No.: 102-539
3. Mazmanian PE, Davis DA. Continuing Medical Education and the Physician as a Learner: Guide to the Evidence. JAMA. 2002; 288(9):1057–1060. [PubMed: 12204068]

4. Davis DA, Taylor-Vaisey A. Translating guidelines into practice: a systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. CMAJ. 1997; 157:408–416. [PubMed: 9275952]
5. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance: a systematic review of the effect of continuing medical education strategies. JAMA. 1995; 274:700–705. [PubMed: 7650822]
6. Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. CMAJ. 1995; 153:1423–1431. [PubMed: 7585368]
7. Carney PA, Sickles EA, Monsees BS, Bassett LW, Brenner RJ, Feig SA, Smith RA, Rosenberg RD, Bogart TA, Browning S, Barry JW, Kelly MM, Tran KA, Miglioretti DL. Identifying Minimally Acceptable Interpretive Performance Criteria for Screening Mammography. Radiology. 2010; 255(2):354–61. [PubMed: 20413750]
8. Warren Burhenne LJ, Smith RA, Tabar L, Dean PB, Perry N, Sickles EA. Mammographic screening: International perspective. Seminars in Roentgenology. 2001; 36(3):187–194. [PubMed: 11475065]
9. Davis, DA.; Lindsay, EA.; Mazmanian, PE. The effectiveness of CME interventions.. In: Davis, DA.; Fox, RD., editors. The Physician as Learner: Linking Research to Practice. American Medical Association; Chicago, Ill: 1994. p. 245-280.
10. Davis D, Thomson MA, Oxman AD, Haynes RB. Evidence for the effectiveness of CME: a review of 50 randomized controlled trials. JAMA. 1992; 268:1111–1117. [PubMed: 1501333]
11. O'Brien, MA.; Freemantle, N.; Oxman, AD., et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. Cochrane Library; Oxford, England: 2002. Update Software
12. Davis DA, Thomson M, Oxman AD, Haynes R. Evidence for the Effectiveness of CME: A Review of 50 Randomized Controlled Trials. JAMA. 1992; 268(9):1111–1117. [PubMed: 1501333]
13. Perry NM. Breast cancer screening--the European experience. International Journal of Fertility & Women's Medicine. 2004; 49(5):228–30. Sep-Oct.
14. Adcock KA. Initiative to Improve Mammogram Interpretation. The Permanente Journal. 2004; 8(2) Spring.
15. Linver MN, Paster S, Rosenberg RD, Key CR, Stidley CA, King WV. Improvements in mammography interpretation skills in a community radiology practice after dedicated courses: 2-year medical audit of 38,633 cases. Radiology. 1992; 184(1):39–43. [PubMed: 1609100]
16. Berg WA, D'Orsi CJ, Jackson VP. Does training in breast imaging reporting and data system (BI-RADS) improve biopsy recommendations of feature analysis agreement with experienced breast imagers at mammography? Radiology. 2002; 224:871–880. [PubMed: 12202727]
17. Carney PA, Geller B, Bogart A, Kerlikowske K, Rosenberg R, Buist D, et al. Association between time spent, confidence and accuracy of screening mammography. AJR Am J Roentgenol. 2012; 198(4):970–8. [PubMed: 22451568]
18. Geller, BM.; Bogart, A.; Carney, PA.; Sickles, EA.; Smith, R.; Monsees, B.; Bassett, L.; Buist, DM.; Kerlikowske, K.; Onega, T.; Yankaskas, B.; Haneuse, S.; Hill, D.; Rosenberg, R.; Wallis, M.; Miglioretti, DL. Educational interventions to improve screening mammography interpretation: a randomized, controlled trial. (In Review)
19. Radiology, ACR. Mammography Case Review. 2012. Available from: http://www.acr.org/mcr
20. Sickles E. The American College of Radiology's Mammography Interpretive Skills Assessment (MISA) Examination. Seminars in Breast Disease. 2003; 6(3):133–9.
21. 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, Chrvala CA, Key CR, Poplack SP, Worden JK, Kessler LG. Breast Cancer Surveillance Consortium: A national mammography screening and outcomes database. American Journal of Roentgenology. 1997; 169:1001–1008. [PubMed: 9308451]
22. Carney PA, Geller BM, Moffett H, Ganger M, Sewell M, Barlow WE, Taplin SH, Sisk C, Ernster VL, Wilke HA, Yankaskas B, Poplack SP, Urban N, West MM, Rosenberg RD, Michael S, Mercurio TD, Ballard-Barbash R. Current Medico-legal and Confidentiality Issues in Large Multi-

center Research Programs. American Journal of Epidemiology. 2000; 152(4):371–378. [PubMed: 10968382]

23. Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, Yankaskas BC, Kerlikowske K, Onega T, Rosenberg RD, Sickles EA, Buist DSM. Variability in Interpretive Performance of Screening Mammography and Radiologist Characteristics Associated with Accuracy. Radiology. 2009; 253:641–651. [PubMed: 19864507]
24. Buist DSM, Anderson ML, Haneuse SJPA, Sickles EA, Smith R, Carney PA, Taplin SH, Rosenberg RD, Geller BM, Onega T, Monsees B, Bassett L, Yankaskas BC, Elmore JG, Kerlikowske K, Miglioretti DL. The influence of volume on interpretive performance of screening mammography in the US. Radiology. 2011; 259:72–84. [PubMed: 21343539]
25. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association. 1952; 47:583–621.
26. Carney PA, Geller BM, Sickles EA, Miglioretti DL, Aiello Bowles E, Abraham L, Feig SA, Brown D, Cook A, Yankaskas BC, Elmore JG. Feasibility and Satisfaction Associated with Using a Tailored Web-based Intervention for Recalibrating Radiologists Thresholds for Conducting Additional Work-up. Academic Radiology. 2011; 18(3):369–376. [PubMed: 21193335]
27. Urban N, Longton GM MS, Crowe AD, Drucker MJ, Lehman CD MD, Peacock S, Lowe KA, Zeliadt SB, Gaul MG. Computer-Assisted Mammography Feedback Program (CAMFP): An Electronic Tool for Continuing Medical Education. Acad Radiol. 2007; 14(9):1036–1042. [PubMed: 17707310]
28. Lewis RS, Bhargavan M, Sunshine JH. Women Radiologists in the United States: Results from the American College of Radiology's 2003 Survey. Radiology. 2007; 242:802–810. [PubMed: 17325067]
29. Patefield WM. Algorithm AS159. An efficient method of generating r x c tables with given row and column totals. Applied Statistics. 1981; 30:91–97.

Pre-intervention test-set data	Radiologist Survey (FAVOR)	BCSC clinical performance data	Pre-intervention test set satisfaction survey	Intervention satisfaction survey	Follow-up test set satisfaction survey	N
●	●	●	●	●	●	43
●	●	●	●	●		5
●	●	●	●		●	8
●	●	●	●			7
●	●	●		●	●	4
●	●	●		●		1
●	●	●			●	1
●	●	●				3
●	●		●	●	●	2
●	●		●		●	2
●	●			●	●	9
●	●			●		4
●	●				●	2
●	●					1
●		●	●	●	●	8
●		●	●	●		1
●		●	●		●	4
●		●	●			3
●		●		●	●	5
●		●			●	1
●		●				2
●			●		●	1
●			●			1
●						1
	●	●				109
	●					6
		●				85
						84*
119	**207**	**290**	**85**	**82**	**90**	**403**

* 84 eligible radiologists (20.8%) provided no data for any part of this study

Figure 1.
Data availability for all 403 eligible invitees

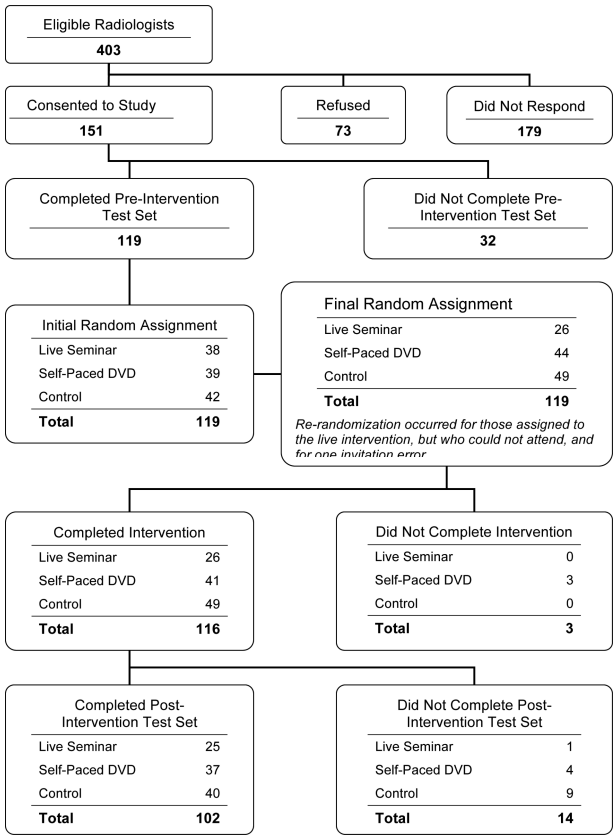

Figure 2.
Flow of Eligible Participants For All Study Components

Table 1

Characteristics of Radiologists among 207 subjects who were eligible to participate in AIM and who completed the FAVOR Survey

	Consenting radiologists who completed all study activities	Consenting radiologists who did not complete all study activities	Non consenting radiologists	P-value
Data Availability:	**N=81/102 (79.4%)**	**N=31/49 (81.6%)**	**N=95/252 (37.7%)**	
Sex				0.03
Male	43 (53.1%)	20 (64.5%)	69 (72.6%)	
Female	38 (46.9%)	11 (35.5%)	26 (27.4%)	
Academic affiliation				0.49
Primary	12 (14.8%)	3 (10%)	7 (7.5%)	
Adjunct	9 (11.1%)	4 (13.3%)	8 (8.6%)	
None	60 (74.1%)	23 (76.7%)	78 (83.9%)	
Missing		1	2	
Fellowship training	12 (14.8%)	5 (16.1%)	7 (7.4%)	0.19
Years interpreting mammograms				0.02
Less than 10	29 (35.8%)	7 (22.6%)	23 (24.2%)	
10 to 20	35 (43.2%)	10 (32.3%)	31 (32.6%)	
More than 20	17 (21%)	14 (45.2%)	41 (43.2%)	
Percent of time spent in breast imaging				0.10
Less than 20%	20 (24.7%)	8 (25.8%)	30 (31.6%)	
20-39%	19 (23.5%)	6 (19.4%)	30 (31.6%)	
40-79%	21 (25.9%)	4 (12.9%)	10 (10.5%)	
80% or more	21 (25.9%)	13 (41.9%)	25 (26.3%)	
Prefer instructor-led CME activities				0.28
Disagree	3 (3.8%)	3 (9.7%)	3 (3.2%)	
Neutral	13 (16.3%)	5 (16.1%)	9 (9.5%)	
Agree	64 (80%)	23 (74.2%)	83 (87.4%)	
Missing	1			
Prefer self-directed CME activities				0.45
Disagree	23 (28.8%)	6 (20.7%)	30 (31.9%)	
Neutral	35 (43.8%)	14 (48.3%)	31 (33%)	
Agree	22 (27.5%)	9 (31%)	33 (35.1%)	
Missing	1	2	1	
Prefer interactive CME activities				0.11
Disagree	3 (3.8%)	3 (10%)	8 (8.6%)	
Neutral	22 (27.5%)	12 (40%)	38 (40.9%)	
Agree	55 (68.8%)	15 (50%)	47 (50.5%)	
Missing	1	1	2	
CME improves my interpretive performance				0.06
Disagree	0 (0%)	3 (9.7%)	1 (1.1%)	

	Consenting radiologists who completed all study activities	Consenting radiologists who did not complete all study activities	Non consenting radiologists	P-value
Data Availability:	**N=81/102 (79.4%)**	**N=31/49 (81.6%)**	**N=95/252 (37.7%)**	
Neutral	14 (17.3%)	3 (9.7%)	16 (16.8%)	
Agree	67 (82.7%)	25 (80.6%)	78 (82.1%)	

All p-values are from Fisher's Exact Test, performed on non-missing observations

Table 2

BCSC performance data in the 5 Years preceding invitation to participate among 290 participants for whom performance data are available

	Consenting Radiologists Who Completed All Study Activities	Consenting Radiologists Who Did Not Complete All Study Activities	Non Consenting Radiologists	P-value
Data Availability:	**N=82/102 (80.4%)**	**N=42/49 (85.7%)**	**N=166/252 (65.9%)**	
Median number of years over the past five during which screening mammogram data was available, median [IQR]	4.34 [2.50, 4.98]	4.91 [2.70, 4.99]	4.96 [4.24, 4.99]	0.004
Number of screening mammograms in the 5 years preceding invitation, n (%)				0.67
1000 or fewer	12 (14.6%)	5 (11.9%)	24 (14.5%)	
1001 - 2000	20 (24.4%)	14 (33.3%)	41 (24.7%)	
2001 - 3000	14 (17.1%)	9 (21.4%)	42 (25.3%)	
More than 3000	36 (43.9%)	14 (33.3%)	59 (35.5%)	
Number of screening mammograms associated with cancer in the 5 years preceding invitation, n (%)				0.88
None	4 (4.9%)	1 (2.4%)	5 (3%)	
1 to 4	12 (14.6%)	6 (14.3%)	24 (14.5%)	
5 to 9	9 (11%)	5 (11.9%)	28 (16.9%)	
10 to 29	30 (36.6%)	19 (45.2%)	56 (33.7%)	
30 or more	27 (32.9%)	11 (26.2%)	53 (31.9%)	
Screening performance measures in the 5 years preceding invitation, median [IQR]				
Sensitivity	86.8 [80, 93.6]	90.0 [82.1, 92.9]	86.8 [77.8, 94.1]	0.49
Specificity	90.7 [87, 94]	91.0 [88.1, 93.3]	91.3 [88.1, 93.8]	0.73
Recall rate	9.7 [6.2, 13.4]	9.5 [7.1, 12.4]	9.0 [6.5, 12.4]	0.68
PPV1	4.1 [2.8, 5.9]	4.7 [3.07, 6.0]	4.3 [3.1, 5.9]	0.57
CDR (cancers per 1000 exams)	4.0 [2.1, 5.6]	4.4 [3.3, 5.4]	3.8 [2.7, 5.3]	0.69

P-values comparing the categorical items are from Fisher's Exact Test. Fisher p-value for the number of screening mammograms associated with cancers was simulated using 10,000 table replicates (27).

P-values for continuous items are from the Kruskall-Wallace test, an extension of the Wilcoxon Rank Sum test which accommodates comparison of more than two groups.

Table 3

Satisfaction with the pre-intervention test set among 85 participants who completed the pre-intervention test set and the corresponding satisfaction survey

	Test set 1: 15 easier cancers	Test set 2: 30 easier cancers	Test set 3: 15 harder cancers	Test set 4: 30 harder cancers	P-value
Data Availability:	**N=20/30 (66.7%)**	**N=25/34 (73.5%)**	**N=18/28 (64.3%)**	**N=22/27 (81.5%)**	
Were the types of abnormal findings on this test set representative of those in your practice?					0.54
Not at all	0 (0%)	0 (0%)	0 (0%)	1 (4.5%)	
Somewhat	5 (25%)	10 (40%)	8 (44.4%)	6 (27.3%)	
Definitely	15 (75%)	15 (60%)	10 (55.6%)	15 (68.2%)	
Do you think the test set was useful for evaluating your skill when interpreting mammography?					0.43
Not at all	3 (15%)	2 (8%)	0 (0%)	1 (4.5%)	
Somewhat	9 (45%)	10 (40%)	5 (27.8%)	7 (31.8%)	
Definitely	8 (40%)	13 (52%)	13 (72.2%)	14 (63.6%)	
Do you think you will change anything in your clinical practice because of this CME exercise?					0.57
Not at all	6 (30%)	4 (16%)	2 (11.1%)	6 (27.3%)	
Somewhat	12 (60%)	15 (60%)	12 (66.7%)	10 (45.5%)	
Definitely	2 (10%)	6 (24%)	4 (22.2%)	6 (27.3%)	
Did you find the feedback on how well you did on the test set compared to your peers helpful to improving your practice?[a]					0.95
Not at all	1 (6.3%)	1 (4.8%)	0 (0%)	2 (12.5%)	
Somewhat	6 (37.5%)	9 (42.9%)	7 (46.7%)	6 (37.5%)	
Definitely	9 (56.3%)	11 (52.4%)	8 (53.3%)	8 (50%)	
Missing				1	
Not assessed	4	4	3	5	
Is this a useful way for you to receive CME mammography credits?					0.56
Not at all	0 (0%)	0 (0%)	0 (0%)	2 (9.1%)	
Somewhat	4 (20%)	8 (32%)	4 (22.2%)	4 (18.2%)	
Definitely	16 (80%)	17 (68%)	14 (77.8%)	16 (72.7%)	

All p-values are from Fisher's Exact Test, performed on non-missing observations

[a] This question was not asked of CME survey respondents from one study site (N=16), all of whom completed the program.

Table 4

Satisfaction With Assigned Intervention among 60 participants in the Live and DVD intervention groups who completed their assigned intervention and completed the corresponding satisfaction survey.

Data Availability:	Live Intervention N=26/26 (100%)	DVD Intervention N= 34/41 (82.9%)	P-value
Were the methods used for instruction appropriate for your learning style?			0.002
Not at all	0 (0%)	2 (5.9%)	
Somewhat	6 (23.1%)	21 (61.8%)	
Definitely	20 (76.9%)	11 (32.4%)	
Do you think the seminar/DVD was useful for evaluating your skill when interpreting mammography?[b]			0.08
Not at all	0 (0%)	5 (14.7%)	
Somewhat	12 (46.2%)	17 (50%)	
Definitely	14 (53.8%)	12 (35.3%)	
Do you think you will change anything in your clinical practice because of this CME exercise?			0.02
Not at all	2 (7.7%)	8 (23.5%)	
Somewhat	11 (42.3%)	20 (58.8%)	
Definitely	13 (50%)	6 (17.6%)	
Did you find the feedback on how well you did answering the questions compared to your peers helpful to improving your practice?			0.28
Not at all	0 (0%)	2 (5.9%)	
Somewhat	8 (30.8%)	15 (44.1%)	
Definitely	18 (69.2%)	17 (50%)	
Is this a useful way for you to receive CME mammography credits?			0.44
Not at all	4 (15.4%)	5 (14.7%)	
Somewhat	5 (19.2%)	12 (35.3%)	
Definitely	17 (65.4%)	17 (50%)	

All p-values are from Fisher's Exact Test, performed on non-missing observations

Table 5

Satisfaction with follow up test set among 88 participants who completed both the follow-up test set and the corresponding satisfaction survey.

Data Availability:	Live Intervention N=23/25 (92.0%)	DVD Intervention N= 34/37 (91.9%)	Control Group N= 31/40 (77.5%)	P-value
Were the types of abnormal findings on this test set representative of those in your practice?				0.31
Not at all	0 (0%)	0 (0%)	0 (0%)	
Somewhat	13 (56.5%)	16 (47.1%)	11 (35.5%)	
Definitely	10 (43.5%)	18 (52.9%)	20 (64.5%)	
Do you think the seminar/DVD was useful for evaluating your skill when interpreting mammography?				0.15
Not at all	0 (0%)	1 (2.9%)	2 (6.5%)	
Somewhat	14 (60.9%)	19 (55.9%)	10 (32.3%)	
Definitely	9 (39.1%)	14 (41.2%)	19 (61.3%)	
Do you think you will change anything in your clinical practice because of this CME exercise?				0.63
Not at all	4 (17.4%)	6 (17.6%)	7 (22.6%)	
Somewhat	14 (60.9%)	20 (58.8%)	13 (41.9%)	
Definitely	5 (21.7%)	8 (23.5%)	11 (35.5%)	
Did you find the feedback on how well you did answering the questions compared to your peers helpful to improving your practice?				0.005
Not at all	1 (4.3%)	3 (8.8%)	0 (0%)	
Somewhat	14 (60.9%)	19 (55.9%)	8 (25.8%)	
Definitely	8 (34.8%)	12 (35.3%)	23 (74.2%)	
Is this a useful way for you to receive CME mammography credits?				0.16
Not at all	0 (0%)	3 (8.8%)	2 (6.5%)	
Somewhat	10 (43.5%)	9 (26.5%)	5 (16.1%)	
Definitely	13 (56.5%)	22 (64.7%)	24 (77.4%)	

All p-values are from Fisher's Exact Test, performed on non-missing observations