

Double Reading of Mammograms by Radiographers: Evidence From the UK Breast Screening Programme

Authors: RL Bennett, RG Blanks, SM Moss

Affiliation: Cancer Screening Evaluation Unit, Institute of Cancer Research
15 Cotswold Road, Sutton, Surrey, SM2 5NG United Kingdom

Keywords: breast neoplasms, mass screening, test sets

Abstract

Background: The UK Breast Screening programme was introduced in 1988. While national protocol recommends that mammograms should be read by a single radiologist or other medically qualified individual, in practise most units use double reading as this has been shown to increase cancer detection rates. Staffing and other resources sometimes do not permit this. To address this problem, radiographers (technologists) in some units have been trained to read mammograms. In the UK, all film readers are encouraged to participate in a self-assessment scheme called PERFORMS, based on a test set. Evidence from this scheme suggests that after allowing for years of experience, radiographers read as well as radiologists. As a result of both recent increases in workload and a shortage of radiologists, and supported by experimental evidence from PERFORMS, some pilot units have introduced double reading by two radiographers, with arbitration by a radiologist or breast clinician.

Objective: To provide evidence from a real-life setting to support this change in reading practise.

Method: An observational study was initiated in 2004, and seven units were identified as pilot sites. A questionnaire was developed and circulated to all screening units to document both their reading practices and the number of years of experience of individual film readers. Using data from the questionnaires and routinely collected performance data, we compared the performance of the pilot units with that of other units.

Results: Provisional results on screening performance measures will be presented.

Test Sets to Evaluate Mammography Interpretative Performance: General Principles and On-Field Experience in a Regional Population Screening

Authors: A Frigerio, L Milanesio, A Ponti, N Segnan

Affiliation: CPO-Piemonte

S. Giovanni Battista Hospital, via S. Francesco da Paola, 31, I-10123 Torino, Italy

Keywords: mammography screening, test sets, film reading performance

Abstract

Background: Achieving breast screening aims of reducing disease-specific mortality heavily depends on the overall quality of the diagnostic procedures, among which film interpretation is of paramount relevance.

Objective: To consider the general difficulties to be overcome in order to create valid mammography test sets, responding to the requisites of being practical to manage, capable of consistently highlighting readers' shortcomings, and useful as teaching tools.

Methods: The method proposed involves utilizing regional data on interval cancer incidence and radiological classification in the setting up of baseline test sets. We also suggest second-line, tailored test sets, adjusted on the individual readers' data on diagnostic performance, as analyzed by the Regional Reference Centre (RRC) in such detail as the individual readers' cancer detection rates and interval cancer rates, correlated with the radiopathologic tumor features. As the RRC for Piedmont, we intend to evaluate these tools in the forthcoming months on a Regional basis.

Results: Results from this on-field evaluation of nine screening districts within a Regional Project shall be discussed.

Conclusion: We believe that a thorough use of both the overall and the individual data on reading performance, as managed by an RRC, may allow optimization of the mammography test set tool, through the development of both baseline sets, primarily aimed to performance evaluation, and tailored test sets, mainly aimed to more effective training. We also believe that many practical difficulties in producing such tools will be substantially overcome in the near future thanks to the availability in digital format of extensive image archives.

Observer Performance in Detection of Breast Cancer Among Hard-Copy Film and Soft-Copy Readings in 3M, 5m-LCD Monitors

Authors: N Ohuchi, T Yamada, A Suzuki, N Uchiyama, A Sato, S Takahashi

Affiliation: Graduate School of Medicine, Tohoku University
1-1, Seiryō-machi, Aoba-ku, Sendai 980-8574, Japan

Keywords: breast cancer screening, digital radiography, data display, ROC curve

Abstract

Background: As the Digital Mammographic Imaging Screening Trial (DMIST) showed the superior ability to detect breast cancer in the subgroups under 50 years old, those with dense breast, and premenopausal women, the ability of digital mammography is acknowledged and digital technique should be more prevalent in breast imaging. In DMIST, CR mammograms were interpreted on hard-copy films. Few studies have compared the diagnostic performance of reading CR mammograms on hard-copy films versus soft-copy.

Purpose: We aimed to compare the ability of soft-copy reading with that of hard-copy reading in the situation close to the screening. We took care in selecting the lesions that were relatively small or screen-detected. This point is different from the previous study, DMIST.

Materials and Methods: We prepared 100 subjects, including 32 patients with surgically proven breast cancer (masses in 12 cases, fine calcifications in 10, and other in 10) and 68 normal controls. All mammograms were obtained by computed radiography (CR) (sampling pitch: 50 μ m). Twelve interpreting doctors independently assessed CR mammograms presented in random order on the three modalities with at least a 4-week interval. Observers rated the images on a 7-point (1 to 7) malignancy scale and on a continuous point malignancy scale (0 to 100). Receiver-operating-characteristics analysis was performed, and average area under the curve (Az) was calculated for each modality. The jackknife method with Bonferroni correction was applied to multireader-multicases analysis.

Results: The average Az of 3M LCD, 5M LCD, and hard-copy film were 0.954, 0.947 and 0.956 on the 7-point scale, and 0.943, 0.923 and 0.944 on the continuous point scale. There was no significant difference among the three modalities on both scales.

Conclusion: Soft-copy reading with 3M- and 5M-LCD was comparable to the reading with hard-copy film to detect breast cancer.

Statistical Methods for Evaluating Mammography Interpretive Performance

Authors: Diana L. Miglioretti, Ph.D.; Sebastien JPA Haneuse, Ph.D.; Melissa Anderson, M.S.

Affiliation: Group Health Center for Health Studies
1730 Minor Avenue, Suite 1600, Seattle, WA 98101

Keywords: test sets, breast cancer detection, screening mammography interpretive performance

Abstract

Background: Extensive variability in screening mammography interpretation exists among radiologists in the United States. Although much research has been conducted toward understanding the role of patient, radiologist, and facility characteristics in explaining this variation, substantial unexplained variability in interpretative performance remains. Recently, interest has turned to evaluating the influence of radiologists' mammographic annual interpretive volume on performance, in part to help inform whether current certification requirements should be changed. Conflicting study findings, however, have defied consensus on whether and how interpretive volume influences performance. These studies shared the common goal of understanding the influence of volume on performance, but they differed in several important ways, including the statistical approaches used to characterize and estimate the associations of interest.

Objective and Methods: We review statistical frameworks commonly used to model interpretive performance, focusing on two broad classes of regression formulations: marginal and conditional models. Both frameworks account for dependence in cluster data, but the interpretations of their parameters differ; hence, the choice of statistical framework may (implicitly) dictate the scientific question being addressed. Additional statistical issues that influence estimation and inference are also discussed, together with their potential impact on the scientific interpretation of the analysis. We illustrate the concepts with data collected by the National Cancer Institute's Breast Cancer Surveillance Consortium and simulations.

Conclusion: It is our hope that careful consideration of the statistical issues relevant to the analysis of mammography interpretative performance will help studies with a common scientific goal, such as determining the influence of interpretive volume on interpretive performance, achieve consensus.

Minimizing Hormone Users' False-Positive Risk at Mammography Screening

Authors: SH Njor, AT Pedersen, W Schwartz, J Hallas, E Lyng

Affiliation: Institute of Public Health, University of Copenhagen
Østre Farimagsgade 5, DK-1014 Copenhagen, Denmark

Keywords: breast cancer, false-positive tests, prior mammograms

Abstract

Purpose: To retrospectively determine impact of comparing current mammograms with prior mammograms on risk of false-positive test for hormone users at screening.

Methods: Data on mammography screening were retrieved for 1993–2005 from Fyn, Denmark. At first screen, two projections were made and at subsequent screens, one projection for fatty and two projections for dense breasts. Until 3 June 2002, 2-year-old mammograms were used for comparison and later 4-year-old mammograms. A positive mammogram not followed by breast cancer was false positive. Prescription drug data were used to identify hormone (HT) use. False-positive risk dependency on age, hormone use, screen number, projection, and prior mammogram was tested with logistic regression.

Results: At first screen, current HT users versus never users had an excess risk of false-positive test of 1.49 (95% confidence interval, CI, 1.22–1.81). At subsequent screens, the risk depended on projections and comparison mammogram. Controlled for number of projections, never users had a lower risk of false-positive test when 4-year-old instead of 2-year-old mammograms were used, 0.79 (95% CI 0.66–0.95). A similar pattern was observed for current HT users, 0.69 (95% CI 0.55–0.86), but not for past users.

Conclusion: Our results indicate that the excess risk of a false-positive test at mammography screening for current HT users compared with never users can be reduced by approximately 40% for women with dense breasts during hormone use, when screening mammograms are compared with 4-year-old instead of 2-year-old mammograms.

Providing Screening Outcome Audit Feedback to Radiologists

Authors: *BM Geller, EJA Bowles

*Affiliation: Health Promotion Research, University of Vermont
429AR4, One South Prospect Street. Elevator C-4426. Burlington, VT 05401-3444

Keywords: mammography, audit feedback

Abstract

Background: Performance feedback can be an effective educational tool to improve future performance. However, there is no previous research on the best way to provide performance feedback to radiologists who perform screening mammography.

Methods: We informally surveyed ICSN members to collect feedback reports from multiple countries and also collected reports from Breast Cancer Surveillance Consortium (BCSC) sites in the United States. We reviewed the literature and existing reports to develop a discussion guide for focus groups and interviews with U.S. radiologists. We asked what data radiologists find useful to improve their screening performance and their preferences for data presentation. We provided examples of various visualized data and created new presentations based on the results from the previous participating radiologists. Qualitative data analysis techniques were used to identify themes and concepts.

Results: We received sample feedback reports from Australia, Canada, Israel, the United Kingdom, and the United States. Most reports consisted of tables with numbers; few contained graphics. The two most important types of data for U.S. radiologists were recall rate and sensitivity. They favored data presented graphically versus numerically. They liked having benchmarks to target and preferred to be compared with other radiologists in their facility, their state, and their country.

Conclusion: Radiologists doing screening mammography preferred visual, graphic data representations. Comparing their outcomes with guidelines and their peers was important to improving their performance. The results will be used to improve the BCSC's outcome audit reports and in an intervention aimed to improve performance.