

Breast cancer screening, diagnostic accuracy and health care policies

Jean-Luc Urbain

§ See related article page 195

The criteria for an effective screening program are well established: the disease should pose a substantial burden, it should have an understood latency period, and there should be acceptable screening tests and therapeutic interventions available. Few would dispute the appropriateness of breast cancer screening (at least for women aged 50–69 years). In most scientific forums, the discussion about breast cancer screening has progressed beyond “whether” to “how.” In this issue (see page 195) Théberge and colleagues seek to maximize the operating characteristics of screening mammography by investigating whether the sensitivity and specificity of the test are associated with the volume of mammograms read by individual radiologists or the volume performed by individual health care facilities.¹ They conclude that radiologists who work in larger facilities and who read more screening mammograms are more likely to have higher breast cancer detection rates and lower false-positive rates.

But before taking these results at face value and reorganizing breast cancer screening systems toward more centralized, high-volume models, we should pause to consider the implications and understand just how breast cancer detection and diagnosis have changed. The conclusions drawn by Théberge and colleagues must be tempered by context and pragmatics. The implications of extracting and extrapolating these conclusions to the performance of decentralized breast cancer screening systems, such as those provided in the United States, where individual facilities and radiologists may encounter relatively small numbers of patients, could result in a gross reduction in the number of radiologists and sites providing screening and, consequently, reduced patient access to breast cancer screening services.

The impact of such changes would not be trivial. Breast cancer is the most common noncutaneous cancer affecting women in North America, with more than 237 000 new cases of invasive disease detected each year.^{2,3} It is second only to lung cancer as a cause of death: in 2004 an estimated 45 000 women will die of breast cancer, as compared with 76 510 women who will die of lung cancer. In the 1970s and 1980s, there was a steady, low increase in the age-adjusted breast cancer mortality, by about 1.5% per year. Since 1991, there has been a sustained reduction in the age-adjusted breast cancer mortality, by about 2% per year.^{2,3} This is likely related to awareness, clinical breast ex-

amination, screening, biopsy, improved therapies and interdependent social factors.

Historically, breast cancer was diagnosed after a woman sought medical attention for a palpable mass or soreness in her breast. This delayed approach is associated with a high risk of death from breast cancer.^{4,5} Over the past 2 decades, it has been shown that mammography can identify breast cancers that are too small to palpate on physical examination. It can also find ductal carcinoma in situ and diagnose small, early-stage breast cancers, including those that have a favourable clinical course. The operating characteristics of screening mammography, influenced by a number of factors, are less than ideal. Although better screening methods are being developed, mammography is currently considered the “gold standard” for breast cancer screening for women over 50.

The sensitivity of mammography in detecting breast cancer depends on the patient’s age, the size and conspicuity of the lesion, the hormone status of the tumour, the overall image quality and the interpretative skills of the radiologist. It also depends on the density of a woman’s breasts, which is affected by her genetic predisposition, hormone status and diet. Retrospective correlations of mammographic findings with population-based cancer registries show that the sensitivity of screening mammography is 70%–90% overall, and ranges from 54%–58% among women under age 40, who tend to have more dense breasts, to 81%–94% among those over 65.⁶ Even though only a small fraction of women (0.1%–0.5%, depending on age) actually have breast cancer when they are screened, the overall specificity of mammography is about 90%.⁶ The number of unnecessary follow-up examinations and procedures illustrates the importance of specificity for breast cancer detection. In the United States, the cost of potentially unwarranted biopsies is more than \$1 billion annually.⁷ In addition, undergoing biopsy for benign disease causes great emotional distress.

A radiologist’s ability to interpret mammograms is critical for the efficacy of screening mammography programs. Previous studies have shown substantial variation in interpretation and reading accuracy among radiologists. Some studies have suggested that sensitivity and specificity are associated with the volume of mammograms read by a radiologist. Others have shown that this factor may not be important.^{8–14} The study by Théberge and colleagues in this

issue¹ shows that the operating characteristics of screening mammography improve with the increasing size of the facility and volume of mammograms read by individual radiologists. This finding is based on a comparative analysis of data of facility and radiologist characteristics associated with 1709 cases of screening-detected breast cancer and 3159 cases of false-positive readings in a 10% random sample of women without screening-detected breast cancer ($n = 30\,560$) obtained from the population-based Quebec Breast Cancer Screening Program. The authors report that the adjusted breast cancer detection rate ratio for facilities performing 4000 or more screenings per year, compared with the rate ratio for facilities performing fewer than 2000, was 1.28 (95% confidence interval [CI] 1.07–1.52). They also report that the adjusted false-positive rate ratio for radiologists who read 1500 or more screening mammograms a year, compared with radiologists who read fewer than 250, was 0.53 (95% CI 0.35–0.79).

These results should be interpreted cautiously. The authors compare the performance of radiologists who interpret at least 1500 screening mammograms per year against a baseline of radiologists who read fewer than 250. Perhaps a more appropriate, clinically relevant, annual baseline volume for individual radiologists would have been 480 mammograms, the minimum annual volume recommended in Canada.¹⁵ Also, the analysis of Th  berge and colleagues fails to account for potential confounders such as the practice of double reading as a quality-control measure. These limitations must be acknowledged to avoid overinterpretation of the influence that mere volume of screening mammograms has on the overall performance of screening mammography programs, particularly since the organization of such programs varies dramatically between countries and continents. In general, countries with socialized medicine (the United Kingdom, Sweden and Canada) have centrally organized systems with high-volume facilities delivering screening mammography services. Other countries, such as the United States, have decentralized systems that are less focused on and are not set up to perform a high volume of procedures and readings. Since the centralized system has resulted in low-cost programs with higher specificity, the attractive information collected from these programs has often been used to implement proscriptive, cost-efficient health care policies for screening mammography.^{12,16}

But comparisons between centralized and decentralized health care systems may be confounded by social, cultural, legal or economic factors that also influence the performance of screening mammography programs. We have to be careful in stating and phrasing our local, regional, provincial and national findings. For example, in the United States, the implementation of centralized screening programs to increase the accuracy of mammography by 10%, by increasing the required volume of services provided by individual sites and radiologists, would result in prohibiting more than 11 000 radiologists from performing this service. The national service capacity would be cut by

50%, and patients would have to travel to specialized, high-volume mammography centres.¹⁶

The most reliable predictor of the success of a screening mammography program is a direct reduction in the rate of advanced-stage breast cancers and an increase in life expectancy among women participating in annual screening mammography. Therefore, the ultimate goal is to decrease the rate of death from breast cancer. To achieve this goal, a responsible health care policy that implements screening mammography programs must find a careful balance between service provision and availability, qualification of radiologists, achievable standards for mammography quality and population demographics.

Dr. Urbain is City-wide Chief of Nuclear Medicine at St. Joseph's Health Centre, London, Ont.

Competing interests: None declared.

References

1. Th  berge I, H  bert-Croteau N, Langlois A, Major D, Brisson J. Volume of screening mammography and performance in the Quebec population-based breast cancer screening program. *CMAJ* 2005;172(2):195-9.
2. Cancer facts and figures 2004. Atlanta: American Cancer Society; 2004.
3. Canadian cancer statistics. Ottawa: Canadian Cancer Society, National Cancer Institute of Canada, Statistics Canada, Provincial/Territorial Cancers Registries, Health Canada, 2004.
4. Newcomb PA, Weiss NS, Storer BE, Scholes D, Young BE, Voigt LF. Breast self-examination in relation to the occurrence of advanced breast cancer. *J Natl Cancer Inst* 1991;83(4):260-5.
5. McDonald S, Saslow D, Alciati MH. Performance and reporting of clinical breast examination: a review of the literature. *CA Cancer J Clin* 2004;54(6):345-61.
6. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology* 2002;225(1):165-75.
7. Burnside E, Belkora J, Esserman L. The impact of alternative practices on the cost and quality of mammographic screening in the United States. *Clin Breast Cancer* 2001;2(2):145-52.
8. Porter PL, El-Bastawissi AY, Mandelson MT, Lin MG, Khalid N, Watney EA, et al. Breast tumor characteristics as predictors of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* 1999;91(23):2020-8.
9. Kerlikowske K, Grady D, Barclay J, Frankel SD, Ominsky SH, Sickles EA, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J Natl Cancer Inst* 1998;90(23):1801-9.
10. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. *N Engl J Med* 1994;331(22):1493-9.
11. Elmore JG, Wells CK, Howard DH, Feinstein AR. The impact of clinical history on mammographic interpretations. *JAMA* 1997;277(1):49-52.
12. Esserman L, Cowley H, Eberle C, Kirkpatrick A, Chang S, Berbaum K, et al. Improving the accuracy of mammography: volume and outcome relationships. *J Natl Cancer Inst* 2002;94(5):369-75.
13. Kan L, Olivotto IA, Warren Burhenne LJ, Sickles EA, Coldman AJ. Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. *Radiology* 2000;215(2):563-7.
14. Beam CA, Conant EF, Sickles EA. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. *J Natl Cancer Inst* 2003;95(4):282-90.
15. Consumer and Clinical Radiation Protection Bureau. *Canadian mammography quality guidelines*. Ottawa: Health Canada; 2002. p. 6.
16. Chirikos TN, French DD, Luther SL. Potential economic effects of volume-outcome relationships in the treatment of three common cancers. *Cancer Control* 2004;11(4):258-64.

Correspondence to: Dr. Jean-Luc Urbain, St. Joseph's Health Centre, 268 Grosvenor St., London ON N6A 4V2; fax 519 646-6403; jeanluc.urbain@sjhc.london.on.ca