Volume of screening mammography and performance in the Quebec population-based Breast Cancer Screening Program

Isabelle Théberge, Nicole Hébert-Croteau, André Langlois, Diane Major, Jacques Brisson

ß See related article page 210

Abstract

Background: In the Quebec Breast Cancer Screening Program (Programme québécois de dépistage du cancer du sein [PQDCS]), radiologists' and facilities' volumes of screening mammography vary considerably. We examined the relation of screening-mammography volume to rates of breast cancer detection and false-positive readings in the PQDCS.

Methods: The study population included 307 314 asymptomatic women aged 50–69 years screened during 1998–2000. Breast cancer detection rates were analyzed by comparing all women with screening-detected breast cancer (n = 1709) and a 10% random sample of those without ($n = 30\,560$). False-positive rates were analyzed by comparing the 3159 women with false-positive readings and the 27 401 others in the 10% random sample. Characteristics of participants, radiologists and facilities were obtained from the PQDCS information system. Data were analyzed by means of logistic regression.

Results: The rate of breast cancer detection appeared to be unrelated to the radiologist's screening-mammography volume but increased with the facility's screening-mammography volume. The breast cancer detection rate ratio for facilities performing 4000 or more screenings per year, compared with those performing fewer than 2000, was 1.28 (95% confidence interval [CI] 1.07–1.52). In contrast, the frequency of false-positive readings was unrelated to the facility's screening volume but was inversely related to the radiologist's screening volume: the rate ratio for readers of 1500 or more screenings per year compared with those reading fewer than 250 was 0.53 (95% CI 0.35–0.79).

Interpretation: Radiologists' and facilities' caseloads showed independent and complementary associations with performance of screening mammography in the PQDCS. Radiologists who worked in larger facilities and read more screening mammograms had higher breast cancer detection rates while maintaining lower false-positive rates.

CMAJ 2005;172(2):195-9

aseload of health care providers and organizations has been linked with performance.¹⁻⁴ Providers with larger volumes of patients or procedures have often been shown to have better outcomes.^{2,5} However, a recent comprehensive review of the literature underlined the methodologic limitations of published studies, especially poor adjustment for case mix and failure to account for char-

acteristics of providers and organizations simultaneously.6

The population-based Quebec Breast Cancer Screening Program (Programme québécois de dépistage du cancer du sein [PQDCS]), launched in 1998, follows the North American standard of a minimum annual reading volume of 480 mammographic examinations (diagnostic and screening combined) for each collaborating radiologist. In addition, facilities in urban areas have to perform at least 4000 diagnostic or screening examinations each year to be eligible for the program. However, there is a large variability in radiologists' and facilities' volumes of screening examinations.

The objective of our study was to assess whether differences in screening volume were associated with rates of breast cancer detection and of false-positive readings. We examined the separate and combined effects of radiologists' and facilities' screening-mammography volumes.

Methods

Between May 1998 and December 2000, 321 985 women had a first screening in the PQDCS. (In this report, we use "screening" to refer to "mammography screening examination"; each screening includes 4 radiographic views; 1 screening corresponds to 1 woman screened.) Of these women, 14 671 were excluded from the study: 13 966 because of breast symptoms (e.g., lump, nipple inversion, nipple discharge), 522 because of prior mastectomy and 183 because of absence of identifying information about the radiologist who read the mammograms. Among the remaining 307 314 asymptomatic women, the mammograms were read as abnormal for 33 938: 1709 (5.6 per 1000) who received a diagnosis of screening-detected in situ or invasive breast cancer within 1 year after the initial abnormal screening result⁹ and 32 229 (10.5%) who did not receive such a diagnosis and therefore had a false-positive initial reading.

All women included in this analysis signed a consent form agreeing to participate in the PQDCS, which includes transmission of their data to a central database for analysis. More than 90% of those screened sign the form. The Commission d'accès à l'information du Québec authorized the use of this database by the Ministère de la Santé et des Services sociaux du Québec (MSSS). In turn, the MSSS mandated analysis of the data by the PQDCS evaluation team.

Information on risk factors for breast cancer is routinely collected with a standardized questionnaire and captured in the province-wide PQDCS information system, which contains the following variables on participants: age, parity, age at first childbirth, menopausal status, age at menopause, history of breast cancer among first-degree relatives, use of hormone replacement therapy, height, weight, mammographic rating of breast density, previous mammography, history of breast reduction or implant, and previous breast aspiration or biopsy.

We took into account several characteristics of the radiologists: sex, year of certification in radiology, number of practices and type (hospital or private clinic), annual volume of screenings, abnormal-recall rate and abnormal-recall rate of colleagues. Sex and year of certification were derived from the Quebec medical directory for the year 1999/2000.¹¹ Other variables were extracted from the PQDCS information system. For a given radiologist, type of prac-

tice and abnormal-recall rate of colleagues were those of the centre where he or she performed the most examinations each year.

We estimated a radiologist's screening volume from the total number of screenings interpreted by the radiologist and the time between the first and last readings, as registered in the PQDCS information system for that individual during the study period. The average monthly volume was then projected over 12 months. For example, a radiologist who interpreted 2000 screenings between Sept. 1, 1998, and Dec. 31, 2000 (28 months), had an estimated annual volume of 857. We estimated the annual screening volume of each facility similarly, using data in the PQDCS information system.

Table 1: Distribution of the 32 269 women included in an analysis of data from the Quebec Breast Cancer Screening Program at initial mammography, May 1998 to December 2000

	Women with	Women without screening-detected breast cancer, no. (%)*	
Characteristic	screening-detected breast cancer, no. (%)* $n = 1.709$	Abnormal mammogram n = 3 159	Normal mammogram n = 27 401
Patients			
Age at initial screening, mean (SD), yr	59.5 (5.9)	57.8 (5.8)	58.3 (5.7)
No. of children, mean (SD)	2.7 (1.5)	2.7 (1.5)	2.7 (1.5)
Age at first childbirth, mean (SD), yr	24.6 (4.7)	23.9 (4.3)	23.9 (4.4)
Age at menopause, mean (SD), yr	47.6 (6.0)	47.1 (6.3)	47.1 (6.2)
Body mass index, mean (SD), kg/m ²	26.6 (5.0)	26.3 (5.1)	26.2 (5.0)
Postmenopausal	1 528 (89.4)	2 667 (84.4)	24 136 (88.1)
Family history of breast cancer	294 (17.2)	425 (13.4)	3 687 (13.4)
Hormone replacement therapy	830 (48.6)	1 642 (52.0)	14 209 (51.8)
Previous mammography	1 405 (82.2)	2 529 (80.1)	24 317 (88.7)
Previous breast aspiration or biopsy	204 (11.9)	388 (12.3)	2 263 (8.2)
Breast reduction	25 (1.5)	99 (3.1)	794 (2.9)
Breast implant	19 (1.1)	45 (1.4)	673 (2.4)
Breast density			
< 25%	390 (22.8)	719 (22.8)	8 896 (32.5)
25%–49%	694 (40.6)	1 205 (38.1)	10 029 (36.6)
50%–75%	513 (30.0)	992 (31.4)	6 765 (24.7)
> 75%	112 (6.6)	243 (7.7)	1 711 (6.2)
Radiologists			
Annual screening-mammography			
volume			
< 250	64 (3.7)	151 (4.8)	965 (3.5)
250–499	324 (19.0)	662 (21.0)	4 851 (17.7)
500–749	442 (25.9)	832 (26.3)	7 529 (27.5)
750–999	375 (21.9)	685 (21.7)	5 600 (20.4)
1000–1249	220 (12.9)	353 (11.2)	3 630 (13.2)
1250–1499	152 (8.9)	296 (9.4)	2 381 (8.7)
≥ 1500	132 (7.7)	180 (5.7)	2 445 (8.9)
Facilities			
Annual screening-mammography volume			
< 2000	324 (19.0)	682 (21.6)	5 620 (20.5)
2000–2999	561 (32.8)	1 061 (33.6)	9 858 (36.0)
3000–3999	412 (24.1)	652 (20.6)	6 984 (25.5)
≥ 4000	412 (24.1)	764 (24.2)	4 939 (18.0)

Note: SD = standard deviation.

^{*}Unless specified otherwise

We analyzed the data for the 307 314 eligible women with a case—control approach. First, we analyzed breast cancer detection rates by comparing the 1709 women with screening-detected breast cancer and a 10% simple random sample of those without such cancer ($n = 30\,560$), as no greater precision was offered by use of a larger sample. Second, we analyzed false-positive rates by comparing the 3159 women with false-positive readings and the 27 401 others in the random sample. In these comparisons, the odds ratios can be interpreted as approximations of the breast cancer detection rate ratios and the false-positive rate ratios. 12,13

We used logistic regression for the multivariate analyses. We evaluated the association of radiologists' screening volume with breast cancer detection and false-positive rates in models, including characteristics of women and radiologists as covariates. Trends in outcomes according to screening volume were tested using the median value within each category of this variable. We used the same strategy to test the association of facilities' screening volumes with breast cancer detection and false-positive rates. Finally, we assessed the independent contribution of the radiologist and the screening facility by fitting models that included both variables simultaneously, after ruling out colinearity. In accordance with the multilevel structure of the data, ¹⁴ we adjusted the confidence intervals (CIs) for the intrareader and intraclinic correlation. We used a 5% level of statistical significance throughout, and all tests were 2-sided.

Results

The distribution of most breast cancer risk factors was similar among the 1709 women with screening-detected breast cancer (1297 invasive and 349 in situ cancers; unknown status in the other 63 cases) and the 30 560 women without this disease (Table 1). For instance, the mean age at

screening was 59 and 58 years, the mean body mass index 27 and 26 kg/m² and the mean age at first childbirth 25 and 24 years, respectively. The proportions of women with a family history of breast cancer, with a previous breast aspiration or biopsy and with dense breasts were higher among the women with than among those without breast cancer.

In contrast, the 3159 women with false-positive mammograms were similar to the 27 401 women with normal mammograms with respect to most breast cancer risk factors (Table 1) except that, compared with the women who had normal readings, fewer of the women with false-positive readings reported previous mammography, more reported a previous breast aspiration or biopsy, and more had dense breasts.

The screening mammograms were interpreted by 275 radiologists working in 68 accredited facilities. Most of the radiologists (73%) were male, and 35% had received their radiology certification recently (between 1990 and 2000). About two-thirds (65%) were working primarily in private clinics. About half (53%) screened at a single facility; 6% screened at 4 or more facilities. The mean annual screening volume was 581 (range 16–2878) per radiologist and 2279 (range 307–7386) per facility.

Breast cancer detection rate ratios and 95% CIs, according to screening volume of either radiologists or facilities, are shown in Table 2. The radiologists' caseload did not seem to influence their ability to detect invasive or in situ breast cancer. The same was true for small invasive cancers (1 cm or less in diameter) (data not shown). By contrast, cancer detection was associated with the number of screen-

Table 2: Adjusted breast cancer detection rate ratios (DRRs),* according to screening-mammography volumes of radiologists and facilities

	No. (%) of	Type of cancer detected; adjusted DRR (95% CI)			
Annual volume	radiologists or facilities	Invasive n = 1297	In situ n = 349	All types n = 1709†	
Radiologists $n = 275$					
< 250	53 (19.3)	1.00	1.00	1.00	
250-499	79 (28.7)	0.93 (0.66-1.33)	1.41 (0.81-2.44)	1.00 (0.73-1.38)	
500-749	74 (26.9)	0.84 (0.59-1.20)	1.29 (0.73-2.26)	0.93 (0.68-1.28)	
750–999	36 (13.1)	0.96 (0.67-1.37)	1.44 (0.80-2.58)	1.05 (0.76–1.45)	
1000-1249	18 (6.6)	1.00 (0.70-1.42)	1.33 (0.71–2.52)	1.05 (0.76–1.46)	
1250-1499	9 (3.3)	0.97 (0.68-1.40)	1.49 (0.80-2.78)	1.06 (0.75-1.48)	
≥ 1500	6 (2.2)	0.87 (0.59-1.29)	1.56 (0.72-3.35)	0.97 (0.67-1.41)	
p value for trend		0.99	0.39	0.71	
Facilities $n = 68$					
< 2000	30 (44.1)	1.00	1.00	1.00	
2000-2999	22 (32.4)	0.92 (0.79-1.08)	1.05 (0.75-1.47)	0.96 (0.82-1.11)	
3000-3999	10 (14.7)	1.09 (0.91-1.30)	1.12 (0.79–1.57)	1.10 (0.94–1.28)	
≥ 4000	6 (8.8)	1.23 (1.03-1.47)	1.51 (1.04-2.19)	1.28 (1.07–1.52)	
p value for trend		0.019	0.028	0.004	

Note: CI = confidence interval.

^{*}Adjusted for participant's age, body mass index, breast density and previous mammography, breast aspiration, biopsy or implant, and for radiologist's abnormal-recall rate, year of certification and number of practices. Radiologists' and facilities' caseloads were also included in models. Thus, DRRs associated with radiologists' screening volumes were adjusted for facilities' volumes and vice versa. †There were also 63 breast cancers of unknown type.

ings performed in the facility. For all types of cancer combined, the adjusted detection rate ratio comparing the facilities performing at least 4000 screenings per year with those performing fewer than 2000 was 1.28 (95% CI 1.07–1.52). The trend according to facility size was highly statistically significant (p = 0.004).

The false-positive rate ratio decreased significantly (*p* value for trend = 0.001) with increasing screening volume of radiologists (Table 3). The false-positive rate ratio

Table 3: Adjusted false-positive rate ratios (FPRRs),* according to screening-mammography volumes of radiologists and facilities

Annual volume	Adjusted FPRR (95% CI)
Radiologists	
< 250	1.00
250-499	1.02 (0.78–1.35)
500-749	0.98 (0.73-1.31)
750–999	1.01 (0.75–1.35)
1000-1249	0.89 (0.63-1.26)
1250-1499	0.93 (0.65–1.33)
≥ 1500	0.53 (0.35-0.79)
p value for trend	0.001
Facilities	
< 2000	1.00
2000-2999	0.91 (0.73–1.15)
3000-3999	0.95 (0.73-1.24)
≥ 4000	1.20 (0.94–1.51)
p value for trend	0.09

^{*}Adjusted for participant's age, body mass index, breast density and previous mammography, breast aspiration, biopsy or implant, and for abnormal-recall rate of radiologist's colleagues and radiologist's year of certification and number of practices. Radiologists' and facilities' caseloads were also included in models. Thus, FPRRs associated with radiologists' screening volumes were adjusted for facilities' volumes and vice versa.

comparing radiologists reading 1500 or more screenings per year with those reading fewer than 250 was 0.53 (95% CI 0.35–0.79). By contrast, the screening volume of facilities was not associated with false-positive readings: the false-positive rate ratio was 1.20 (95% CI 0.94–1.51) for those performing 4000 or more screenings per year compared with those performing fewer than 2000.

Table 4 shows the combined association of radiologists' and facilities' caseloads with performance. Radiologists who worked in facilities performing a greater number of screenings per year had higher detection rates than those who worked in facilities performing fewer, and this was true for all radiologists working in high-volume facilities, irrespective of their individual screening volume. In contrast, the false-positive rates decreased with increasing radiologist caseload, and this trend was clearer among those who worked in larger facilities.

Interpretation

Our results support the notion that radiologists' and facilities' screening volumes are independently associated with performance and that their effects may be complementary. In our analysis, radiologists who read larger numbers of screening mammograms and worked in facilities performing larger numbers of screenings tended to have higher breast cancer detection rates while maintaining lower false-positive rates than radiologists who performed fewer readings and worked in facilities performing fewer screenings.

A few studies^{15–19} have examined the association of radiologists' interpretation volume with their performance. Four of them^{15–18} were conducted as experiments; since performance in a test environment appears to be unrelated to performance in actual practice, ^{20,21} the relevance of the results of these

Table 4: Adjusted breast cancer DRRs and FPRRs,* according to screening-mammography volumes of both radiologists and facilities

Annual volume	No. of radiologists $n = 275$	Breast cancer detection		False-positive readings	
		No. of cases/controls† n = 1709/30 560	Adjusted DRR (95% CI)	No. of cases/controls‡ n = 3159/27 401	Adjusted FPRR (95% CI)
Facilities with volume $< 3000, n = 52$					
Radiologists					
1-499	101	259/4669	1.00	568/4092	1.00
500-999	57	351/7261	0.99 (0.84–1.17)	710/6551	1.06 (0.85–1.31)
≥ 1000	18	275/5291	1.08 (0.92-1.28)	465/4835	0.85 (0.65-1.12)
Facilities with volume ≥ 3000 , $n = 16$ Radiologists					
1–499	31	129/1960	1.20 (0.98–1.48)	243/1715	1.29 (1.03–1.61)
500–999	53	466/7385	1.20 (1.02–1.43)	807/6578	1.12 (0.92–1.36)
≥ 1000	15	229/3994	1.25 (1.03–1.52)	366/3630	0.92 (0.68–1.24)

^{*}DRRs were adjusted for participant's age, body mass index, breast density and previous mammography, breast aspiration, biopsy or implant, and for radiologist's abnormal-recall rate, year of certification and number of practices. FPRs were adjusted for participant's age, body mass index, breast density and previous mammography, breast aspiration, biopsy or implant and for abnormal-recall rate of radiologist's colleagues and radiologist's year of certification and number of practices.

[†]Cases are women with screening-detected breast cancer. Controls are women free of breast cancer with normal or abnormal mammograms

[‡]Cases are women free of breast cancer with abnormal mammograms. Controls are women free of breast cancer with normal mammograms.

studies to our purposes is questionable. Only Kan and colleagues19 evaluated the relation of radiologists' reading volume with performance in an operating population-based breast cancer screening program, like ours. They showed a trend, although not significant, to improved cancer detection with greater reading volume. They also showed a U-shaped relation between abnormal-recall rate (a proxy for falsepositive rate) and radiologists' screening-mammography volume. In that study, conducted in British Columbia, the smallest and largest categories of annual screening volume were < 2000 and 4000-5199; in our analysis the corresponding volumes were < 250 and 1500–2878. In addition, Kan and colleagues could adjust only for age and history of previous screening, and they did not study the relation of facility screening volume to performance.

Our data suggest an association between a centre's screening volume and its cancer detection rate but not its false-positive rate. Blanks and associates²² observed a significant increase in invasive breast cancer detection rates and a decrease in abnormal-recall rates with increasing facility screening volume. Yankaskas and coworkers23 also observed lower recall rates with increasing facility screening volume but saw no change in sensitivity. Neither group took into account radiologists' reading volume.

Our study had some limitations. First, there were few radiologists or facilities with large numbers of screenings, and this limited our ability to compare our results with those in settings with higher radiologist or facility screening volumes. Second, our study covered only the first 2 years of the PQDCS, which might not be representative of current functioning. Third, only screening-mammography volumes of radiologists and facilities were available to us, whereas accreditation criteria are based on both screening and diagnostic experience. Finally, other important determinants of performance, such as double reading, participation in teaching or research, daily quality-control procedures within facilities and specific training of the radiologists in the interpretation of screening mammograms, could not be taken into account.

In conclusion, radiologists' and facilities' screening volumes appear to have independent and complementary influences on performance as measured by rates of breast cancer detection and false-positive readings. Although the average crude breast cancer detection rates met the minimum Quebec standard of 5 cancers per 10007 in all categories of volume shown for radiologists and facilities, the overall performance of screening mammography seems to be maximized when screenings are performed in larger centres and when, in these centres, mammograms are read by radiologists who interpret a large volume of films.

This article has been peer reviewed.

From Direction des systèmes de soins et services, Institut national de santé publique du Québec, Sainte-Foy, Que. (all authors), Centre de recherche, Hôpital Charles LeMoyne, Greenfield Park, Que. (Hébert-Croteau), and Unité de recherche en santé des populations, Hôpital du Saint-Sacrement du Centre hospitalier affilié universitaire de Québec, Québec, Que. (Brisson)

Competing interests: None declared.

Contributors: All of the authors participated in the planning and design of the study, the data interpretation and critical revision of the manuscript; all approved the version to be published. Isabelle Théberge was responsible for statistical analysis

Acknowledgements: We thank Drs. Michel-Pierre Dufresne, Sylvie Groleau, Michel Petitclerc, Robert Pronovost and Guy Roy for their helpful comments. We also thank members of the working group for the evaluation of the Quebec Breast Cancer Screening Program for their valuable suggestions.

References

- 1. Hillner BE, Smith TJ, Desch CE. Hospital and physician volume or specialization and outcomes in cancer treatment: importance in quality of cancer care. 7 Clin Oncol 2000;18:2327-40.
- 2. Birkmeyer JD, Siewers AE, Finlayson EV, Stukel TA, Lucas FL, Batista I, et al. Hospital volume and surgical mortality in the United States. N Engl J Med
- Hébert-Croteau N, Brisson J, Pineault R. Review of organizational factors related to care offered to women with breast cancer. Epidemiol Rev 2000;22:228-38.
- Hannan EL. The relation between volume and outcome in health care. NEngl J Med 1999;340:1677-9.
- Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. Surgeon volume and operative mortality in the United States. N Engl J Med 2003;349:2117-27.
- Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. Ann Intern Med 2002;137:511-20.
- 7. Association canadienne des radiologistes. Programme d'agrément en mammographie. Ville Saint-Laurent (QC): l'Association; 2000. p. 5
- Ministère de la santé et des services sociaux. Programme québécois de dépistage du cancer du sein. Cadre de référence. Québec: Gouvernement du Québec; 1996. p. 14.
- Théberge I, Major D, Langlois A, Brisson J. Validation de stratégies pour obtenir le taux de détection du cancer, la valeur prédictive positive, la proportion des cancers in situ, la proportion des cancers infiltrants de petite taille et la proportion des cancers infiltrants sans envahissement ganglionnaire dans le cadre des données fournies par le Programme québécois de dépistage du cancer du sein (PQDCS). Sainte-Foy (QC): Institut national de santé publique du Québec; 2003. p. 17.
- Ministère de la santé et des services sociaux. Programme québécois de dépistage du cancer du sein. Rapport d'activités des années 1998 et 1999. Québec: Gouvernement du Québec; 2001. p. 15.
- Collège des médecins du Québec. Annuaire médical 1999-2000. Montréal: le
- Cornfield J. A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. J Natl Cancer Inst 1951;11:1269-75.
- 13. Miettinen O. Estimability and estimation in case-referent studies. Am J Epidemiol 1976;103:226-35.
- Diez-Roux AV. Multilevel analysis in public health research. Annu Rev Public Health 2000;21:171-92
- 15. Beam CA, Conant EF, Sickles EA. Association of volume and volumeindependent factors with accuracy in screening mammogram interpretation. 3 Natl Cancer Inst 2003;95:282-90.
- Elmore JG, Wells CK, Howard DH. Does diagnostic accuracy in mammography depend on radiologists' experience? 7 Womens Health 1998;7:443-9.
- Ciatto S, Ambrogetti D, Catarzi S, Morrone D, Rosselli DT. Proficiency test for screening mammography: results for 117 volunteer Italian radiologists. 3 Med Screen 1999;6:149-51
- 18. Esserman L, Cowley H, Eberle C, Kirkpatrick A, Chang S, Berbaum K, et al. Improving the accuracy of mammography: volume and outcome relationships. 7 Natl Cancer Inst 2002;94:369-75
- 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;215:563-7.
- Rutter CM, Taplin S. Assessing mammographers' accuracy. A comparison of
- clinical and test performance. J Clin Epidemiol 2000;53:443-50.

 Elmore JG, Miglioretti DL, Carney PA. Does practice make perfect when interpreting mammography? Part II. J Natl Cancer Inst 2003;95:250-2.

 Blanks RG, Bennett RL, Wallis MG, Moss SM. Does individual programme
- size affect screening performance? Results from the United Kingdom NHS breast screening programme. J Med Screen 2002;9:11-4.
- Yankaskas BC, Cleveland RJ, Schell MJ, Kozar R. Association of recall rates with sensitivity and positive predictive values of screening mammography. Am J Roentgenol 2001;177:543-9.

Correspondence to: Dr. Jacques Brisson, Direction des systèmes de soins et services, Institut national de santé publique du Québec, 945, ave. Wolfe, 5^e étage, Sainte-Foy QC G1V 5B3; fax 418 682-7949; jacques.brisson@uresp.ulaval.ca