Re: International Variation in Screening Mammography Interpretations in Community-Based Programs

In the article by Elmore et al. (1), the authors conclude that abnormal recall rates in North American screening programs appear to be higher than those in other countries without evident benefit in cancer detection rates. As a representation of screening mammography in North American countries outside the United States, the authors cite findings from 11,824 mammography screening examinations conducted during the first 15 months of a single organized program in British Columbia, Canada, between 1988 and 1989. Since that time, program-based mammography screening has grown in Canada, with published outcome data (as of 1999–2000) available for 3,476,433 screened women from 10 organized screening programs (2–5). For Elmore et al. to make inferences regarding the state of mammography screening in “North America” on the basis of their limited data generalizes their findings too broadly—that is, the data by Elmore et al. is too limited in terms of coverage, amount (i.e., data were included from only one in 10 programs from Canada and only seven programs from the United States), and timeliness to be considered representative of current screening practices in North America as a whole.

Indicators of screening program performance provide useful information for program evaluation; however, these indicators should be considered within the context of the screening program as a whole. In addition, inference about differences in program performance requires more comprehensive analysis of quantitative and qualitative differences between programs than those offered by Elmore et al. (1). For example, Elmore et al. do not fully consider that quantitative relationships between abnormal recall rates, positive predictive value, and cancer detection rates alone do not provide a meaningful indication of program performance. Further information about interval cancer detection rates and the underlying breast cancer incidence rates are required for such interpretation. Elmore et al. also do not consider qualitative differences between programs that impact upon the unwanted effects of screening as indicated by low positive predictive values. In Canada, follow-up for the majority of abnormal patient recalls consists of noninvasive procedures. For example, in 1999–2000, imaging procedures, including diagnostic mammographic examination and/or ultrasound, were the only assessment required for nearly three-quarters of women with a screening abnormality (4). Consequently, the positive predictive value of biopsy performed in Canada presently meets or exceeds both Canadian and international standards (6).

Hence, if the comparison by Elmore et al. of positive predictive values of biopsy performed had distinguished between the types of diagnostic assessment, including the type of biopsy performed (i.e., fine-needle aspiration versus core biopsy versus open biopsy), the relative burden of diagnostic assessment may have been more apparent.

International comparisons of mammography screening programs can provide valuable information to aid in optimizing their screening performance when rigorous approaches are applied to contrast and combine results from different countries. The Performance Parameters Evaluation Group of the International Breast Cancer Screening Network—a consortium of 25 countries—is systematically assessing the feasibility of making international comparisons of screening program performance (see http://appliedresearch.cancer.gov/ibsn/research/activities.html). Their approach includes the standardization of performance measure definitions, data collection methods, and the formulas for calculating program performance and the consideration of confounding contextual factors, such as the inclusion of clinical breast examination during a screening examination, rates of non-program screening in the program’s target population, and the completeness of cancer registration. In the absence of such an approach, international comparison of screening program performance deserves cautious interpretation and provides limited information for the refinement of screening program policy.

Jay Onysko
Christina M. Bancej
Carolyn P. Pim

REFERENCES


NOTES

Affiliations of authors: Chronic Disease Prevention Division, Centre for Chronic Disease Prevention and Control, Health Canada, Ottawa, Ontario, Canada (JO, CMB); Database Management Committee, Canadian Breast Cancer Screening Initiative, Calgary, Alberta, Canada (CPP).

Correspondence to: Jay Onysko, MA, Chronic Disease Prevention Division, Centre for Chronic Disease Prevention and Control, Health Canada, 120 Colonnade Rd., P.L. 6701B, Ottawa K1A 1B4, Canada (e-mail: jay.onysko@hc-sc.gc.ca).

DOI: 10.1093/jnci/djh040

RESPONSE

Our literature search, as outlined in the “Methods” section of our recent article (1), reviewed MEDLINE peer-reviewed publications up to mid-2002 that included three search terms: “mammography,” “mass screening,” and “biopsy”. The references noted by Onysko et al. in their correspondence were not identified by our search because they were either published in 2003 (i.e., were not in the peer-reviewed literature at the time of our search) or did not use the specified key terms.
Screening practices in Canadian breast cancer screening mammography programs may differ from those in the United States in several ways. Indeed, the medical malpractice environment, fiscal incentives, health care structure, and breast cancer screening recommendations differ between the two countries. We strongly suspect that our findings reflect differences in screening performance mainly between U.S. screening mammography programs and screening results from other countries.

Our findings are consistent with another recently published article (2), in which screening mammography performance was compared between the United States and the United Kingdom. In that study, the recall rate and negative open-surgical biopsy rates were twice as high in United States settings as they were in the United Kingdom, whereas the cancer detection rates were similar in both countries.

We agree with Onysko et al. about the need for standardization of international comparisons of screening program performance. However, the data required for meaningful comparison of screening performance are, unfortunately, not always available in the existing published literature (for example, the type of diagnostic assessment performed or definitions used to calculate outcomes). The conceptual model and brief discussion of the potential reasons for variability among published studies of screening mammography [Table 3 in (1)] may be helpful in making comparisons of screening program performance in future studies.

JOANN G. ELMORE
DAVID F. RANSOHOFF

REFERENCES


NOTES

Affiliations of authors: Department of Medicine, University of Washington, Seattle (JGE); Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill (DFR).

Correspondence to: Joann G. Elmore, MD, MPH, Department of Medicine, Division of General Internal Medicine, Harborview Medical Center, 325 Ninth Ave., Box 359780, Seattle, WA 98104-2499 (e-mail: jelmore@u.washington.edu).

DOI: 10.1093/jnci/djh041