Increasing Value by Increasing Volume: Call for Changes in US Breast Cancer Screening Practices

Christoph I. Lee, Joann G. Elmore

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

The optimal effectiveness of screening mammography is reliant on the widely variable interpretive accuracy of a human reader: the radiologist (1). It seems logical that radiologists with more clinical experience, such as those routinely interpreting high volumes of mammograms, are more accurate in their interpretations than radiologists with less experience. Following this logic, most countries with breast cancer screening programs have minimum volume requirements among their interpreting radiologists to achieve and maintain high levels of screening accuracy. Surprisingly, compared with most European countries, the United States has the lowest requirements for radiologists to be eligible to interpret screening mammography, requiring interpretation of only 960 screening exams over a 2-year period (2–5). This volume threshold is dwarfed by European Commission guidelines that recommend a minimum volume of 5000 mammograms per year for interpreting radiologists (5).

As most of us have experienced, skills fade over time when they aren’t being used. This phenomenon is often referred to as being “out of practice.” The association between physician volume and patient outcomes has been extensively studied in various fields of health care. In general, a higher volume of procedures among individual physicians is associated with better outcomes across a wide variety of conditions (6).

In this issue of the Journal, Théberge et al. offer one of the more convincing analyses to date regarding the association between radiologists’ interpretive volume and accuracy in interpreting mammography (7). Their comprehensive analysis includes all radiologists and imaging facilities in the Quebec breast cancer screening program over a several-year period. This study avoids inherent selection bias and controls for variations of interpretive volume over time among individual radiologists, two problems that have plagued prior studies (8–11). They found that radiologists who consistently interpreted less than 500 mammograms annually had lower accuracy compared with those who consistently interpreted 500 or more mammograms annually. Moreover, increases in interpretive volume (up to approximately 3000 mammograms annually) were associated with persistent improvements in radiologists’ accuracy. These findings corroborate prior studies linking higher interpretive volumes with decreases in false-positive rates without much effect on sensitivity (8–10,12).

Although high-volume practice on a regular basis is needed to keep skills and abilities honed, how one practices may also be important. Radiologists interpreting just 1000 screening mammograms a year may only see five exams with breast cancer, while they call back another 100 women for additional evaluation when no cancer is present (false positives). Unfortunately, these radiologists may not receive feedback on the outcome of all women whose screening mammograms they call back for additional testing. With such low exposure to breast cancer cases and without immediate and effective feedback on which abnormal screening exams resulted in the true detection of cancer, radiologists confined to screening mammography will not effectively refine their skills. In contrast, radiologists involved in the entire care continuum from screening through the diagnostic work-up, including performing image-guided biopsies, obtain valuable and immediate personal feedback that may improve their ability to distinguish suspicious from benign findings.

Interestingly, Théberge et al. found that radiologists with experience in diagnostic mammography have higher accuracy than those interpreting only screening mammography. This finding is similar to a recent US-based analysis showing that adding a minimal diagnostic interpretation volume requirement in addition to increasing the minimum screening volume requirements would result in marked reductions in false-positive results without adversely affecting cancer detection (13). False-positive screening mammograms are both costly and anxiety provoking and were a major driver of the controversial US Preventive Services Task Force revised recommendations in 2009 recommending informed decision making for women in their 40s and biannual screening among women aged 50 years and older.

So why is the US volume requirement for radiologists interpreting mammograms lower than that of other countries, especially in light of the growing body of evidence showing that higher volume is associated with higher accuracy? The arbitrary minimum US threshold was originally chosen to assure adequate access to mammography by not eliminating the large pool of radiologists in rural settings with low annual mammography volumes. However, both the minimum threshold set two decades ago and the arguments against increasing it now seem glaringly outdated. The issue of maintaining adequate patient access to screening mammography can now be addressed by the widespread adoption of digital mammography and modern telediagnosis practices. Women can now obtain screening mammograms in a rural area on a mobile mammography van and have the images sent electronically for immediate interpretation to a high-volume subspecialty-trained radiologist at a diagnostic center hundreds of miles away.
Continuing to maintain low-volume and sporadic readers in the pool of radiologists interpreting screening mammography may also not be needed or desirable from a workforce standpoint. The impact of doubling the current annual volume requirements, from our current level of approximately 500 to 1000 exams each year, on the US workforce capacity may be relatively minimal. In an analysis of a national sample of US radiology practices, only 17% of interpreting radiologists had an annual screening volume of less than 1000 mammograms (14). Removing this small proportion of radiologists, responsible for only 5% of the total screening volume, from the interpretation pool may actually be a welcome change for these individuals; compared with higher-volume radiologists, US radiologists who sporadically interpret mammography have reported greater malpractice fears, more insecurities associated with interpretation, and express a preference to not interpret mammography altogether (15).

Renewed focus with regards to increasing the minimum mammographic interpretive volume requirements in the United States is further warranted by the Affordable Care Act and the mandate for higher quality health care at decreased costs. Screening mammography costs the United States at least $7 billion annually in direct costs (13,16). A large proportion of the direct costs are due to false-positive findings at screening, resulting in unnecessary diagnostic evaluations and interventional procedures. In addition to direct costs, false-positive findings have indirect costs, including patient anxiety, unnecessary morbidity, and lost time. Efforts to reduce false positives resulting from breast cancer screening programs are therefore warranted.

In medicine, we are constantly trying to improve processes, with the main focus on ultimately improving the well-being of patients. As we review the greater picture of breast cancer screening practices in the United States, it is time for us to acknowledge that clinical skills require diligent upkeep and practice and, to ensure quality health care for women, our clinical practices need to change. After all, practice makes perfect, or at least, it makes us better.

References

Affiliations of authors: Department of Radiology (CIL) and Department of Medicine (JGE), University of Washington School of Medicine, Seattle WA; Department of Health Services (CIL) and Department of Epidemiology (JGE), University of Washington School of Public Health, Seattle WA.