How Best to Determine the Mortality Benefit From Screening Mammography: Dueling Results and Methodologies From Canada

Russell P. Harris

In this issue of the Journal, Coldman et al. report results from the Pan-Canadian Study of Mammography Screening and Mortality from Breast Cancer, an analysis of observational data from seven of 12 provincial screening programs from 1990 to 2009 (1). They find that “participation in mammography screening programs in Canada was associated with substantially reduced breast cancer mortality,” with a relative risk reduction of 40%. Interestingly, five of the seven provinces involved in this study were also sites for the Canadian Breast Screening Study (CNBSS), a randomized controlled trial (RCT) that recently reported, after 25 years of follow-up (1980 to 2005), no reduction in breast cancer mortality from screening mammography (2). If we are to determine the mortality benefit from current breast cancer screening, and to monitor it over time, we need to understand the lessons from the fascinating contrast presented by these two studies.

In addition to the common setting, many aspects of the two studies were similar. In the Pan-Canadian provincial study, women ages 40 (five provinces started at 50) to 79 years (five provinces stopped at 69) self-referred for screening; in the CNBSS, women ages 40 to 59 years volunteered to be randomized. Mammography in the two studies appears to have been similar: two-view, single-read, analog mammography. CNBSS screened annually for five years; the provincial screening programs screened biennially (two provinces screened women aged 40–49 years annually). CNBSS included clinical breast examination (CBE) in some screening arms; some provinces used CBE and some didn’t. Referral to treatment in both studies was determined by primary care physicians.

The study methodologies, however, were quite different. The CNBSS was an RCT, assuring similarity of compared groups by randomization. The provincial study, on the other hand, is an observational study that pulls together data from each of the seven screening programs, provincial cancer registries, and provincial mortality databases in an attempt to compare observed breast cancer mortality among “participants” with expected mortality if they had not been screened. They derive their estimate of expected mortality from a multistep indirect approach using data from “nonparticipants.” As the authors are aware, this indirect approach is prone to error at each step: adjusting for out-migration, deconstructing total provincial breast cancer incidence rates into specific rates for participants and for nonparticipants, deconstructing breast cancer survival rates in the same way, reconstructing incidence and survival data with appropriate demographic and other adjustments to estimate breast cancer deaths for the participant group in the absence of screening from the data on nonparticipants. But screening participants and nonparticipants differ from each other in multiple ways, some known and some unknown. Calculating expected mortality for participants using data from nonparticipants is hazardous; our confidence in these estimates must be low. To try to deal with the problem of differences between participants and nonparticipants, the Pan-Canada authors use a small substudy with another complex analysis that, for many reasons, is ultimately unconvincing.

It would not be surprising for the two studies to differ based only on methods. A classic example comes from yet another Canadian study—the 2004 report of 11-year follow-up from the Quebec Prospective Randomized Controlled Trial of prostate cancer screening (3). The unfortunate analysis of this RCT found a 62% reduction in prostate cancer mortality from screening. This result came from ignoring randomization and treating the study population as an observational cohort, comparing mortality in all screened men (participants) against all unscreened men (nonparticipants), regardless of randomized group. When one uses an intention-to-screen analysis, however, comparing prostate cancer mortality in the group randomized to screening invitation with the group randomized to receive no screening invitation, there is no difference in mortality. The lesson is clear—people who participate in screening programs differ from people who do not, and the differences include determinants of the outcome.

Thus, in the duel between RCT and observational studies, there is no contest: RCTs simply provide better evidence about mortality in screening programs. Yet the Pan-Canada authors make important points about the need for observational evidence. RCTs cannot easily be used to monitor the effects of screening over time. But that is exactly what is needed. With improving treatment (a potential confounder that is inadequately dealt with in the Pan-Canada study), it is possible that at some point—whether now or in the future—breast cancer screening will become less effective and the harms of screening may outweigh the benefits (4,5). We need to know when that time has come. In addition, newer screening tests are being developed that are promoted as finding more breast cancers; some then extrapolate without further evidence to conclude that this increased detection would lead to reduced mortality (6). Some of these screening tests are currently being implemented (7). We need ongoing studies to monitor and compare the effects of...
different screening programs—including both benefits and harms. Although not ideal, as we have seen, observational studies are more suited for this task than RCTs.

But there is also a duel ongoing among observational studies. One example is the Pan-Canada authors’ use of “incidence-based mortality” to count breast cancer deaths in screened vs unscreened groups. This involves counting only the subset of breast cancer deaths “which would be affected by screening.” Although this sounds reasonable, it turns out to be very difficult and controversial in practice (8). As the Pan-Canada authors note, studies that use this approach usually find a much higher mortality reduction than studies that do not. It is not at all clear which approach is best. Another methodological issue in observational studies is whether we are more interested in estimating the effect of screening on a population vs the effect on individuals who adhere carefully to a regular screening schedule over many years (4). Observational studies focused on individuals adjust for participation in screening, while studies interested in population-level effects do not. Such adjustments often change the answer. There are still further unsettled issues in observational studies, such as how best to adjust for underlying secular trends in incidence and for the effects of treatment (4), and how to interpret change (or no change) in the incidence of late-stage disease at diagnosis (8).

From the 1970s up to the recent CNBSS report, the mortality benefit of breast cancer screening has been assessed primarily by RCTs. Given changes in breast cancer risk factors, improvements in treatment, and changes in screening technology, the older RCTs now are less and less credible for this assessment. Quite appropriately, more recent studies have assessed the breast cancer screening benefit by observational studies, often comparing populations either screened or unscreened (4). These studies have the advantage of large numbers and can examine changes over time, comparing areas with one screening strategy against another. But now we are witnessing dueling results even from different observational studies, one study finding large benefits from screening (9) while others find little benefit (10). These studies vary in both investigators and methodologies. Yet we are stuck with observational studies to help us answer the ongoing questions about screening: is screening still effective? Is one screening technology or strategy better than another? Are there populations that need more or less screening? What we need is not merely more observational studies; what we need is more trustworthy observational studies.

In the end, this editorial is a plea for an end to the duel among observational studies, a plea for better observational methods and for a plan for cooperation in answering current questions about breast cancer screening. Specifically, we need:

- to convene an international panel of impartial methodologists to develop standards for trustworthy observational studies—their design, conduct, data requirements, and analysis—and how to assist screening programs to use these standards in implementing ongoing monitoring systems; and to develop an international plan for cooperating in using data from the monitoring systems to compare the benefits and harms of breast cancer screening among programs that differ in population risk factors or screening technology.

Otherwise, we can expect to be buffeted by continual duels between studies with results determined as much by methodology as by the true effects of screening.

References

5. Elmore JG, Harris RP. The harms and benefits of modern screening mammography. BMJ. In press.

Note

The author has no conflicts of interest to declare.

Affiliation of author: Research Center for Excellence in Clinical Preventive Services, University of North Carolina, Chapel Hill, NC.