

Point: Mammography screening—sticking to the science

M.J. Yaffe PhD*

BACKGROUND

The concept underlying screening is that, by detecting potentially lethal cancers in a population at an earlier point than when those cancers would surface clinically, earlier and less harsh treatment can be given, reducing both mortality and morbidity. The efficacy of screening women between the ages of 40 and 69 for breast cancer has been demonstrated in individual randomized controlled trials and meta-analyses, with overall mortality reductions varying between 19% and 31%^{1–4}. Furthermore, observational studies of screening as delivered in more than 20 organized programs has shown mammography screening to be effective, with participating populations being associated with mortality reductions of 40% or higher compared with mortality in nonparticipating populations over the age range of 40–74 years^{5,6}.

Of necessity, a program that aims to reduce the incidence rate of advanced disease through earlier detection requires that many asymptomatic individuals be screened to detect the relatively small number of cancers in a population. That requirement causes screening, even for a relatively common disease such as breast cancer, to be a rather inefficient process. Most women—about 93%–95% of them—will receive a negative result, and their only benefit will be the reassurance that they do not have breast cancer. That inefficiency could be one of the reasons that some physicians and researchers are not supportive of screening. Wouldn't it would be better to image only those with cancer or only those who have major risk factors? Unfortunately, if cancer is to be found earlier, it is not acceptable to wait until women become symptomatic. And most women who develop breast cancer carry only the risk factors of being female and more than 40 years of age.

THE CANADIAN NATIONAL BREAST SCREENING STUDY

The Canadian National Breast Screening Study (CNBSS) used mammography or clinical physical examination, or both, to screen women between the years 1980 and 1985, and in 1992, the investigators first published the study's failure to observe a benefit from mammography screening^{7,8}. In his November 2014 article⁹, Dr. Steven Narod defended the CNBSS, which has been heavily criticized by other scientists on the basis of methodologic flaws and poor mammography quality^{10–12}.

In 1993, I was co-author on an article, which Narod unfortunately failed to cite, that laid out the criticisms of the CNBSS without casting aspersions about the motives or underlying reasons for the deficiencies in that trial¹⁰. Because of the onslaught against screening mammography presented in the paper by Narod and the editorials by Baum¹³ and Foulkes¹⁴, I thought that it would be useful to revisit the problems that caused the CNBSS to stand apart from the other randomized controlled trials of breast cancer screening in being the only trial that concluded with a higher breast cancer death rate in the group invited to screening compared with the control group. Those problems are likely the factors that prevented the CNBSS from demonstrating a mortality reduction from screening in 1992, when the first publication of its results appeared^{7,8}. Because the decisions and basic conditions underlying the problems could not subsequently be modified, it is not surprising that a mortality reduction did not emerge upon the second publication of the trial results in 2002^{15,16} or in the 25-year follow-up¹⁷. The largest of the problems is probably the evidence that some women with poor-prognosis cancers were not randomly entered into the two arms of each trial, but there are others that should not be ignored.

Statistical Power

The CNBSS is actually not a single trial, but two separate trials, asking two different research questions. In CNBSS1, 50,000 women 40–49 years of age were to be randomized to receive either annual mammography plus clinical breast examination (CBE) by a trained nurse for 5 years or an initial CBE followed by “usual care” over the subsequent 4 years. Here, the research question was “Does screening in the 40–49 age group contribute to reduced mortality from breast cancer?”

In CNBSS2, the research question was “Does the addition of mammography screening increase the mortality reduction provided by CBE alone?” In this trial, approximately 40,000 women 50–59 years of age were to be randomized to 5 annual examinations which were either CBE or CBE plus mammography.

Each of the CNBSS studies was powered to be able to detect a mortality reduction of 40% or greater^{7,8}. We are beginning to see 40% mortality reductions in the current screening era^{5,6}, but effects of that size certainly did not emerge from any of the CNBSS era trials, and the CNBSS trials did not have adequate power to detect a smaller mortality reduction if it existed. Although the two trials had been kept

separate through the first two publications, the authors of the 2014 publication¹⁷ chose, for reasons that were not explained in their article, to combine the results of these trials that asked two different research questions, inflating the size from 40,000 and 50,000 for the original trials to 90,000, possibly giving readers an overly optimistic impression of the study power.

Randomization

Randomization in the CNBSS used a decentralized open-book method in which slots were to be assigned randomly, allocating women to the study or the control arm in each of the two trials. But, in the initial (prevalence) round of screening, women received a CBE *before* their names were entered, and there would therefore be some knowledge at the screening site of palpable abnormalities before the official registration of the participant occurred. Boyd *et al.*^{10,18} observed that, in the data presented in the first CNBSS1 publication, of 24 women with poor-prognosis breast cancers discovered in the prevalence round, 19 were assigned to the study (mammography) arm and only 5 to the control arm. Of the 19 cancers in women assigned to the study arm, 17 were palpable, and so those cancers were not simply found earlier because of the lead time afforded by mammography. The probability of this assignment occurring by chance was estimated by Boyd at 0.0033. That estimate is a strong indication of a failure in randomization, where, technically, for any variable of interest, the distribution in the study group and the control group should be approximately 50:50.

The excess of breast cancer deaths that would occur from even a small imbalance in the assignment of women who entered the study with advanced cancers would certainly shift the study results away from a demonstration of mortality reduction. In fact, Miller *et al.*¹⁷ provide evidence of that shift. In 1995, Tarone¹⁹ suggested that the CNBSS data be analyzed excluding the deaths that occurred from cancers detected in the prevalence round. When that analysis was finally done by the CNBSS authors in 2014, the hazard ratio associated with screening fell 15 percentage points to 0.9 (95% confidence interval: 0.69 to 1.16) from 1.05 (95% confidence interval: 0.85 to 1.3). Without speculating on the exact nature of the cause of that fall, the change is compelling evidence of a randomization problem. Furthermore, the authors reported that the breast cancer mortality hazard ratio associated with the mammography arm was 1.47 for cancers detected in the prevalence screen, compared with 0.9 in subsequent screens¹⁷, further supporting the argument that the initial randomization was not balanced, with the imbalance conferring a bias against observing a benefit from screening.

Image Quality and Sensitivity

The ability to achieve a mortality reduction from screening is highly associated with the ability to achieve a reduction in the rate of advanced cancers that appear in the screened population²⁰. Although the randomized trials that demonstrated mortality reductions observed reductions greater than 20% in advanced cancer rates (node-positive or larger than 2 cm in diameter), the CNBSS did not. Further, the mean size of cancers detected was only about 2 mm smaller in the mammography arm than in the control arm. When

imaging fails to detect breast cancer smaller than that which can be detected with palpation, the implication is that either the quality of the mammographic images, or the quality of the interpretation by the radiologist, or both, are suboptimal.

In screening, the ability to reduce the incidence rate of advanced cancers comes from the detection and treatment of cancers earlier in their natural history. Earlier detection requires that the images be of high quality with respect to breast positioning and technical characteristics (contrast, resolution, signal-to-noise ratio) and that the interpretations be performed by highly skilled radiologists. A series of several prominent breast radiologists (Wende Logan-Young, Stephen Feig, Edward Sickles, Daniel Kopans, Laszlo Tabar) brought in by the CNBSS during the screening period assessed the quality of the mammography. Those radiologists consistently expressed concern with the quality, considering the images to be below the standards of the time^{11,21}.

As a consultant to the study, initially asked to provide input on issues related to radiation safety, I also noted limitations—in some cases severe limitations—with the technology available for imaging at the clinics and hospital facilities that performed the imaging for the CNBSS²². Some of the mammography systems were old and lacked features that are essential for producing mammograms of consistently high quality. In some cases, I noted that far-from-optimal exposure techniques were being used. It is likely that those problems contributed to the inability in the CNBSS to reduce the incidence of advanced cancers and to the failure of CNBSS to observe mortality reductions from screening similar to those in the other randomized trials.

Crossover or Contamination

In randomized trials of cancer drug therapy, patients in the study arm are offered a new agent thought to have to potential to improve outcome, while those in the control arm receive the standard treatment. Normally, patients are motivated to be assigned to the study arm, believing that the new drug could help them. Meanwhile, those in the control group do not have access to the drug outside the trial. Although some patients drop out of the trial, there are fewer opportunities for “crossover” than exist in a randomized screening trial in which study participants have a low probability of having cancer and might, for example, fear radiation and not attend screening, while those in the control group can easily access mammography outside of the trial. In the CNBSS, 26% of the women 40–49 years of age in the non-mammography arm underwent mammography at least once during the 5-year screening period; in the older group, 17% had at least 1 mammography exam. Although the outside use of mammography is not, *per se*, a flaw in the trial, if the analysis performs no correction for the effect, any real mortality reduction will be markedly underestimated.

Overdetection

Often incorrectly called “overdiagnosis,” the term “overdetection” refers to cancers detected by screening, which, because of indolence would never otherwise have surfaced in the individual’s lifetime. Although overdetection almost certainly occurs at some level, there has

been much controversy recently over the magnitude of the phenomenon. In principle, a randomized trial with a long follow-up period would be an excellent platform for an evaluation of overdiagnosis. Miller *et al.*¹⁷ reported an overdiagnosis rate of 22%. Although the 2014 paper was to be a 25-year follow-up, the overdiagnosis calculation was, for reasons not explained, done at the 15-year point. If the CNBSS data are observed at 25 years, the estimated level of overdiagnosis is much lower. The total number of invasive cancers was 3250 in the mammography arms and 3133 in the control arm, an imbalance of only 117 (3.7%). Because the degree of screening uptake by women in the two arms of each study after the 5-year screening period is not known, the reliability of either of those estimates is likely to be low^{23–25}.

SUMMARY

There is strong evidence, much of it provided in the 25-year update¹⁷, to explain why the CNBSS is indeed an outlier among randomized trials in failing to demonstrate a mortality reduction associated with an invitation to mammography screening. Although Narod's loyalty to his mentor is commendable, I believe that if he were to study this evidence carefully, he would understand why the CNBSS results should not be used to influence screening policy. To the extent that his article and the articles by Baum¹³ and Foulkes¹⁴ in the same issue of *Current Oncology* use the CNBSS results as the foundation of their arguments discounting the value of mammography screening, while ignoring the large body of evidence supporting the benefits of mammography, they do an injustice to women for whom such programs represent one of the few proven interventions that can reduce their risks of death from breast cancer and of the morbidity associated with treatment of advanced disease.

ACKNOWLEDGMENTS

Some of the points raised here were originally published in Boyd *et al.*, 1993¹⁰.

CONFLICT OF INTEREST DISCLOSURES

I have read and understood *Current Oncology's* policy on disclosing conflicts of interest, and I declare the following interests: From 1990 to 2015, I was a consultant on image quality and radiation safety to the Ontario Breast Screening Program. Much of my professional activity focuses on reducing mortality and morbidity from cancer through improved imaging. My laboratory has a research collaboration with GE Healthcare on digital breast tomosynthesis. I am a founder and shareholder in Matakina Technology, a company that develops software for measuring mammographic density.

AUTHOR AFFILIATIONS

* Physical Sciences/Imaging Research, Sunnybrook Research Institute, University of Toronto, and Ontario Institute for Cancer Research, Toronto, ON.

REFERENCES

1. Independent U.K. Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;380:1778–86.
2. Nelson HD, Tyne K, Naik A, *et al.* Screening for breast cancer: systematic evidence review update for the U.S. Preventive Services Task Force. *Ann Intern Med* 2009;151:727–37;W237–42.
3. Hendrick RE, Smith RA, Rutledge JH 3rd, Smart CR. Benefit of screening mammography in women aged 40–49: a new meta-analysis of randomized controlled trials. *J Natl Cancer Inst Monogr* 1997;87–92.
4. Duffy SW, Yen AMF, Chen TH, *et al.* Long-term benefits of breast screening. *Breast Cancer Manag* 2012;1:31–8.
5. Coldman A, Phillips N. Population studies of the effectiveness of mammographic screening. *Prev Med* 2011;53:115–17.
6. Coldman A, Phillips N, Wilson C, *et al.* Pan-Canadian study of mammography screening and mortality from breast cancer. *J Natl Cancer Inst* 2014;106:dju261.
7. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *CMAJ* 1992;147:1459–76. [Erratum in: *CMAJ* 1993;148:718]
8. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *CMAJ* 1992;147:1477–88. [Erratum in: *CMAJ* 1993;148:718]
9. Narod SA. Reflections on screening mammography and the early detection of breast cancer. *Curr Oncol* 2014;21:210–14.
10. Boyd NF, Jong RA, Yaffe MJ, Trichler D, Lockwood G, Zylak CJ. A critical appraisal of the Canadian National Breast Cancer Screening Study. *Radiology* 1993;189:661–3.
11. Kopans DB. The Canadian Screening Program: a different perspective. *AJR Am J Roentgenol* 1990;155:748–9.
12. Burhenne LJ, Burhenne HJ. The Canadian National Breast Screening Study: a Canadian critique. *AJR Am J Roentgenol* 1993;161:761–3.
13. Baum M. Re: Reflections on screening mammography and the early detection of breast cancer. *Curr Oncol* 2014;21:215–16.
14. Foulkes WD. Screening mammography: the turning of the tide? *Curr Oncol* 2014;21:205–7.
15. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med* 2002;137:305–12.
16. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years. *J Natl Cancer Inst* 2000;92:1490–9.
17. Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ* 2014;348:g366.
18. Boyd NF. The review of randomization in the Canadian National Breast Screening Study. Is the debate over? *CMAJ* 1997;156:207–9.
19. Tarone RE. The excess of patients with advanced breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. *Cancer* 1995;75:997–1003.
20. Morrison AS. Intermediate determinants of mortality in the evaluation of screening. *Int J Epidemiol* 1991;20:642–50.
21. Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. *AJR Am J Roentgenol* 1993;161:755–60.
22. Yaffe MJ. Correction: Canada Study [letter]. *J Natl Cancer Inst* 1993;85:94.
23. Smith RA. Counterpoint: overdiagnosis in breast cancer screening. *J Am Coll Radiol* 2014;11:648–52.
24. Miller AB. Re: “Counterpoint: overdiagnosis in breast cancer screening.” *J Am Coll Radiol* 2014;11:923.
25. Smith RA. Re: “Counterpoint: overdiagnosis in breast cancer screening” [author's reply]. *J Am Coll Radiol* 2014;11:923–4.