COMMENTARY

Reflections on screening mammography and the early detection of breast cancer

A Countercurrents Series[a]
with S.A. Narod MD

> A little learning is a dangerous thing.
> — Alexander Pope, *An Essay on Criticism*

In the stormy aftermath of the recent publication of results from the 25-year Canadian National Breast Screening Study (NBSS)[1], various opinions questioning the validity of the study's results have been expressed[2–7]. I was a latecomer to the study. In 2005, I was charged with oversight of the final record linkage and the statistical analysis and interpretation of the final data set. Dr. Anthony Miller has been my mentor since 1987. Our first joint paper, on screening for cervical cancer, was published in 1991[8]. I chose not to respond to individual criticisms, but instead to collect my thoughts and to try to explain why the study authors saw no benefit from screening.

Most of the criticism from the radiology community focuses on issues of study design (which they claim was inadequate) and on the quality of the mammography (which they also claim was inadequate). Cancer survivors bolster those criticisms with testimonials and appeals to common sense. Supporters of the study are drawn from the public health community, and they tend to focus on overdiagnosis and health economics.

The report at issue is not the first emerging from the NBSS. Earlier reports[9,10] were criticized for not having allowed adequate follow-up time. But the 25-year results resemble the early results, and the authors are no longer criticized for premature disclosure. None of the first-generation critics have acknowledged the consistency; instead, they look elsewhere and point out other weaknesses. They claim that high-risk women were assigned to the mammography arm in violation of the principle of randomization. In his bestseller *The Emperor of All Maladies,* Siddhartha Mukherjee says, *as a matter of fact,* that high-risk women were

assigned surreptitiously to the mammography arm, which explains the lack of observed benefit[11].

The most recent NBSS report[1] tallied the breast cancers that occurred in each of the two study arms after the screening period ended (that is, between years 6 and 25), counting 2584 cancers in the screening arm and 2609 cancers in the control arm. If the screening arm had been enriched for women at "high risk," that enrichment must have been performed in a peculiar fashion, using only risk factors that have a transient effect. Perhaps Dr. Mukherjee would care to explain what those factors were. It follows that the excess of cancers seen in the screening period (years 1–5: 666 vs. 524) was a result of early diagnosis and not from stacking the deck.

In any case, compelling evidence against the criticism of assignment of high-risk women to the screening arm is provided in the most recent analysis[1], and that criticism is no longer raised (although no one has retracted or apologized). Instead, critics now insist that many women with palpable lesions were sent directly to the screening arm by duplicitous research assistants. There is no reason to believe that such actions (which would involve a national conspiracy of dozens of coordinators who spoke two official languages) were taken, but even if they had been, the study and its conclusions would not necessarily be invalidated. Even if all the women with prevalent cancers had been shunted to the screening arm, the situation could still be remedied by ignoring all cancers found at the first screening round (prevalent cancers) and focusing instead on the incident cancers. Such a strategy is not uncommon in screening studies. In the NBSS, no woman had the opportunity to "cross the floor" from one study arm to the other after initial assignment. Therefore, if we exclude all prevalent cases from the analysis and focus on women with no cancer at study entry, we can re-evaluate the benefit of mammography thereafter. The hazard ratio for death from breast cancers detected in screening rounds 2–5 was 0.90 (95% confidence interval: 0.69 to 1.16; *p* = 0.40).

But what about crossover? It is claimed that a certain proportion of the women in the control arm—perhaps as high as 20%—opted for screening off-study, in particular after the screening period was over. That crossover will, some say, eclipse a benefit of screening that might otherwise have ensued. That is, the benefit of mammography (which might well have been substantial) was nullified by a subcohort of independently-minded women who went for mammography at the end of the 5 years. That speculation is fanciful, but if true, should be welcomed, because it can now be said to a patient who, at age 40, requests a mammogram, that there is no hurry; she can come back in 5 years for a mammogram and achieve the same net benefit. And when she comes back at age 45, she can be reprieved again until age 50.

Crossover is a form of contamination that results in misclassification of the exposed and unexposed groups. In a trial, it will tend to bias the result toward the null. The best way to avoid misclassification is to randomize the patients after they agree to participate—as the NBSS did. In contrast, in the Swedish two-county trial (discussed in more detail a little later in this article), the subjects were randomized by intention-to-treat—that is, by whether they received or did not receive an invitation to mammography[12–15]. Of the 78,085 women in Sweden who were offered screening, 69,645 accepted and 8440 declined. In effect, then, 8440 women in the Swedish study were *de facto* misclassified (versus an undisclosed number of hypothetical crossers-over in the Canadian study). The proponents of the Swedish study do not see that misclassification as a shortcoming, but instead use it to buoy their argument in favour of screening. They say that if everybody invited for screening came for screening, then the protective effect would have been more profound. In the Swedish study, all women in the control group were offered a screening test after the screening period ended (a reasonable thing to do); but those authors were not criticized for "contaminating" their study.

The second issue raised concerns the quality of the mammography. After all, the NBSS tests were completed 30 years ago using 30-year-old technology. I still wonder how things might have been done differently. Mammography screening identified 212 women with breast cancer who would otherwise have been missed. They had cancers that were, on average, 1.4 cm in size, with 67% being node-negative. The survival of those women was very good. At the end of the study period, 170 women with a nonpalpable mammography-detected breast cancer were alive or had died of other causes. How many of those lives did screening save? Fifty? Twenty-five? Ten? Unfortunately, all we can say is that the number was too few to be noticed. If a significant number of those 170 lives had, in fact, been saved, surely the difference between study arms would have been noticeable. Breast cancer deaths numbered 180 in the mammography group and 171 in the control group. Perhaps

some of the survivors believe that their lives were saved. They might perhaps have written a letter to the editor of their local newspaper extolling the virtues of mammography. But 42 women with a nonpalpable mammography-detected cancer died (none of whom has written a letter to the editor).

I am also among the authors of several publications on the benefits of screening by magnetic resonance imaging (MRI) in high-risk women[16–18]. Those studies were greeted as successes, given that they demonstrated how, with the use of MRI, breast cancers could be downstaged. Those studies were accepted by the radiology community as being supportive of screening. Whether MRI reduces mortality has not yet been shown. I cannot predict whether MRI screening will be effective in reducing mortality 10 years down the line, but I fully expect that if a mortality benefit fails to materialize, the studies will be criticized for using 30-year-old equipment and a poor study design.

Much of the criticism of the NBSS has come from Drs. Daniel Kopans and László Tabár, and fellow travellers such as Siddhartha Mukherjee and Patrick Borgen[2–7,11]. They use the Swedish two-county trial as evidence of a good study that supports the use of mammography and quote a 30% reduction in mortality. Naturally, they do not criticize their canonical study, but it is time to take a closer look.

In the NBSS, women were randomized on an individual basis after they had attended the study centre. The result was two groups of equal size and 100% compliance with the first screen. In Sweden, the two counties were divided into 19 geographic strata that were then divided into either 2 blocks (Östergötland) or 3 blocks (Kopparberg). The resulting 45 blocks were randomized, and women in more than half the blocks were sent a letter of invitation to screening. Of the 59% of women who received an invitation, 89% came for the first screen and 83% came for the second screen[14].

The Canadian women were offered 5 mammograms 1 year apart. The Swedish women were offered mammograms every 2 years (ages 40–49) or every 3 years (ages 50–74) for up to 8 years. They underwent fewer screens (Table I). The cancers detected by mammography in Canada were similar in size to those detected in Sweden (Table I), but the size of the cancers occurring in the control group were very different. Those comparisons suggest that physical examinations or breast cancer awareness (or both) were important contributors to the size of cancers detected in Canada. A diminution of cancer mortality would not be expected to be associated with a 0.2 cm mean difference in tumour size, but might be expected with a net reduction of 0.7 cm in size[19]. Of the cancers detected in the screening arm of the Canadian trial, 68% were palpable. That fact has been a source of criticism. But a physical examination was not conducted as part of the screening protocol in Sweden, and the comparable number of palpable tumours was not given. Therefore, given the much longer mean time between screening

TABLE I   A comparison of key parameters in the Canadian National Breast Screening Study (NBSS) and the Swedish two-county trial

Variable	NBSS	Two-county trial
Participants (*n* women)	89,835	133,065
Randomization	Individual (89,835)	Geographic cluster (45 units)
Proportion in screening group (%)	50	58
Compliance with first screen (%)	100	89
Mammogram average (*n*)[a]	4.2	2.3[b]
Screening interval (months)	12	24–36
Cancers detected [*n* (%)]		
During screening period	1190 (1.32)	2468 (1.85)
After screening period	6383	NR
Palpable lesions (%)	68.2	NR
Size of cancers (cm)		
Screened group	1.9	1.9[c]
Control group	2.1	2.6[c]
Case fatality rate (%)		
Screened	27.0	24.6
Unscreened	32.6	35.2

[a]   Among women screened.
[b]   Estimated from Tabàr *et al.,* 1992, Table 4[14].
[c]   Estimated from Tabàr *et al.,* 1992, Table 10[14].
NR = not reported.

visits in Sweden, and the high proportion of women in the screening arm that were never screened, I estimate that between 70% and 80% of the cancers in the mammography arm in Sweden would have been palpable and could have been detected by physical examination—had it been done. The fact that the relevant number is not given is a critical lapse. Suppose, for the sake of argument, that 100% of the cancers detected in the screening arm in Sweden were in fact palpable (not a gross exaggeration). What then would be the point of mammographic screening? And if that number (the palpable fraction) is not available, how can the results be judged? Neither the Swedish nor the Canadian trial can exclude the possibility that the benefit from invitation to mammography might have been restricted to women with palpable cancers.

The Canadian study reports the number of cancers detected in the follow-up period after the end of the screening period and the number of subsequent deaths from breast cancer. From year 6 to year 25, 2584 incident cancers occurred in the screening group, resulting in 298 deaths (11.5%), and 2609 incident cancers occurred in the control group, resulting in 321 deaths (12.3%). Those data are important because they confirm that, in the absence of screening, the cancer incidence and mortality are equal in the study groups. Where are the comparable numbers for the Swedish study? Again, they are not given. But in

looking at the extraordinary Figure 1 from the most recent report of the Swedish study[12], the mortality curves are seen to continue to separate at *25 to 29 years after the initiation of screening,* and long since screening had stopped.

Tabár and colleagues ask readers to believe that the benefits of mammography are everlasting (or at least for 20 years beyond the end of screening). They make that claim despite having no surety about whether the deaths from breast cancer in years 25–29 were the result of cancers diagnosed during the screening period or diagnosed after screening had stopped. They claim that most of the deaths from breast cancers diagnosed in the control arm occurred more than 10 years after diagnosis. Thus, the reader is asked to accept that a mean of 2.3 mammograms obtained in year 1–7 are more likely than a baseline imbalance in breast cancer risk to lead to a reduction in breast cancer mortality of 30% in years 25–29!

The incidence and mortality rates corresponding to cancers that were diagnosed after the screening trial was stopped are unavailable. Seeing the survival curves corresponding to cases detected in the screened and unscreened cohorts would be interesting. In the NBSS, most cancer deaths occurred, as expected, within 10 years from diagnosis[1]. When the NBSS was challenged as to having achieved an even balance in the study groups, the authors provided the relevant data. The Swedish authors should do the same. Patrick Borgen has stated that the NBSS is the "worst clinical trial ever done"[5]—an extraordinary statement. Either he has devoted his life to poring over medical tracts with the zeal of a Talmudic scholar, or he is speaking nonsense. But refuting his claim is easy: it takes merely the time required to read the Swedish papers.

Once the facts are accepted (that screening mammography fails to do what it was intended to do, and that overdiagnosis is real and substantial), then the most interesting questions can begin to be addressed. Did the NBSS fail because mammography is not a sufficiently sensitive imaging technique? Or has the screening community been working under false premises?

Consider sensitivity. Proponents of mammography say that the technique is currently better than it was in the 1980s, largely because it is more sensitive. (Specificity is also important, but is not at issue here.) They argue that "the more sensitive, the better." The earlier a cancer can be identified and managed, the better. The smaller, the better. But those contentions generate an interesting paradox. Consider a woman with a small early-stage breast cancer. The recommendation is that this woman be followed with annual bilateral mammography for 5 or more years to identify recurrences and contralateral cancers[20]. That recommendation is based on the knowledge that the risk of contralateral cancer is between 0.5% and 0.8% annually[21] and that a diagnosis of contralateral cancer is associated with an increase in mortality

from breast cancer[22]. (It has not been shown that screening for contralateral cancer reduces mortality.) But MRI is a much more sensitive screening tool than mammography, and by using MRI in that setting, a small contralateral breast cancer can be identified in 4% of women with newly-diagnosed breast cancer[23]. And yet routine MRI of the contralateral breast is not recommended, because it has not been shown to improve survival. Instead, the recommendation for follow-up with annual mammography continues. The paradox is this: If 8 years' worth of incident breast cancers can be identified in one shot, why bother to pick them up in dribs and drabs? The MRI-detected occult lesions are understood not to be clinically meaningful because they do not adversely affect mortality (overdiagnosis); however, if a similar lesion were to be found as a primary cancer in the ipsilateral breast, the radiologists insist that it is clinically meaningful. Once the paradigm that an increase in sensitivity increases overdiagnosis is accepted (that is, not all lesions are *clinically meaningful*), then it is the responsibility of clinicians to try to determine the ideal level of sensitivity.

The NBSS has been berated for working with 30-year-old machinery, but I think that the greater problem is that clinicians are still working under 30-year-old assumptions. How much is really known about the relationship between size and survival? How confident is our community about early detection? It is universally accepted that tumour size and survival are inversely related for women diagnosed with palpable breast cancer[24]. That understanding is the rationale for early detection by mammography or other means. But it does not *logically* follow that a decrease in tumour size will necessarily lead to a decrease in mortality.

Consider two analogous situations. First, among women with breast cancer who experience a local recurrence, the strongest predictor of death is a short time from diagnosis to local recurrence[25]. However, that finding does not imply that a further shortening of the time from diagnosis to recurrence through intensive imaging would worsen survival. Second, studies of children with neuroblastoma noted that the children diagnosed in the first year of life experienced much better survival than those diagnosed thereafter[26]. That observation encouraged physicians to consider that screening for neuroblastoma by measuring urinary metabolites would increase the proportion of children diagnosed in the first year and thereby reduce mortality. The resulting clinical trial unfortunately found no benefit[27]. Neuroblastoma with a favorable prognosis is detectable by screening, but those cases are associated with a very high rate of spontaneous regression or maturation of the neuroblastoma into benign ganglioneuroma. Very few cases of neuroblastoma detected by screening have unfavourable biologic features such as N-Myc amplification[28].

The relationship between breast cancer size and survival is not fixed, and the slope of the curve that defines the relationship varies according to the stage and pathologic features of the breast cancer[24]. The strongest relationship is seen with large cancers and node-positive cancers[29]. The relationship is attenuated among women with triple-negative cancers, with HER2 (human epidermal growth factor receptor 2)–positive cancers, and with *BRCA1*-positive cancers[19,30]. Size does not predict mortality well for women with nonpalpable cancers[29]. Is it possible that there are additional categories wherein the size–survival relationship does not hold, and that eventually every woman with breast cancer will be able to be assigned to one of those categories? If more specific categorization were to be possible, then there would be no expectation of benefit from early detection—through mammography or any other means. In statistical terms, the question is "Are there variables n_1, n_2, n_3, ... n_x, such that, after adjusting for n_1, n_2, n_3, ... n_x in a follow-up study, size is no longer predictive of survival?" For example, in a study of 5423 women with cancers of less than 2.0 cm, tumour size was not predictive of survival after adjustment for grade, hormone receptor status, and HER2 expression[30]. Those data suggest that, as the mean size of breast cancers in a population diminishes, further reductions in size can achieve only marginally less benefit. The lesson of mammography should be used to rethink the fundamentals of breast cancer and its natural history so that planning can commence for the experiments and clinical studies that will lead to better outcomes in the future.

CONFLICT OF INTEREST DISCLOSURES

The author has no financial conflicts of interest to disclose.

REFERENCES

1. 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.
2. Tabár L, Chen THH. We do not want to go back to the dark ages of breast cancer screening [Web letter]. *BMJ Response* London, U.K.: BMJ; 2014. [Available online at: http://www.bmj.com/content/348/bmj.g366/rr/686788; cited August 2, 2014]
3. Tabár L. A failure from the beginning [Web letter]. *BMJ Response* London, U.K.: BMJ; 2014. [Available online at: http://www.bmj.com/content/348/bmj.g366/rr/687126; cited August 2, 2014]
4. Borgen PI. The real tragedy .... [Web letter]. *BMJ Response* London, U.K.: BMJ; 2014. [Available online at http://www.bmj.com/content/348/bmj.g366/rr/686662; cited August 2, 2014]

5. Fand Incollingo B. Canadian Breast Screening Study Reignites Controversy Over Mammograms [Web article]. Philadelphia, PA: OncLive; 2014. [Available online at: http://www.onclive.com/conference-coverage/mbcc-2014/Canadian-Breast-Screening-Study-Reignites-Controversy-Over-Mammograms; cited cited August 2, 2014]

6. Kopans DB. Having been one of the experts called on in 1990 to review ... [Web letter]. *BMJ Response* London, U.K.: BMJ; 2014. [Available online at: http://www.bmj.com/content/348/bmj.g366/rr/686292; cited August 2, 2014]

7. Kopans DB. Arguments against mammography screening continue to be based on faulty science. *Oncologist* 2014;19:107–12.

8. Narod SA, Thompson DW, Jain M, Wall C, Green LM, Miller AB. Dysplasia and the natural history of cervical cancer: early results of the Toronto Cohort Study. *Eur J Cancer* 1991;27:1411–16.

9. 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.

10. 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.

11. Mukherjee S. *The Emperor of All Maladies: A Biography of Cancer.* New York, NY: Scribner; 2010.

12. Tabár L, Vitak B, Chen TH, *et al.* Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology* 2011;260:658–63.

13. Tabár L, Fagerberg CJ, Gad A, *et al.* Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet* 1985;1:829–32.

14. Tabár L, Fagerberg G, Duffy SW, Day NE, Gad A, Gröntoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am* 1992;30:187–210.

15. Tabàr L, Duffy SW, Vitak B, Chen HH, Prevost TC. The natural history of breast carcinoma: what have we learned from screening? *Cancer* 1999;86:449–62.

16. Warner E, Plewes DB, Hill KA, *et al.* Surveillance of *BRCA1* and *BRCA2* mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA* 2004;292:1317–25.

17. Warner E, Hill K, Causer P, *et al.* Prospective study of breast cancer incidence in women with a *BRCA1* or *BRCA2* mutation under surveillance with and without magnetic resonance imaging. *J Clin Oncol* 2011;29:1664–9.

18. Passaperuma K, Warner E, Causer PA, *et al.* Long-term results of screening with magnetic resonance imaging in women with *BRCA* mutations. *Br J Cancer* 2012;107:24–30.

19. Narod SA. Tumour size predicts long-term survival among women with lymph node-positive breast cancer. *Curr Oncol* 2012;19:249–53.

20. Grunfeld E, Dhesy–Thind S, Levine M on behalf of the Steering Committee on Clinical Practice Guidelines for the Care and Treatment of Breast Cancer. Clinical practice guidelines for the care and treatment of breast cancer: follow-up after treatment for breast cancer (summary of the 2005 update). *CMAJ* 2005;172:1319–20.

21. Narod SA. Bilateral breast cancers. *Nat Rev Clin Oncol* 2014;11:157–66.

22. Schaapveld M, Visser O, Louwman WJ, *et al.* The impact of adjuvant therapy on contralateral breast cancer risk and the prognostic significance of contralateral breast cancer: a population based study in the Netherlands. *Breast Cancer Res Treat* 2008;110:189–97.

23. King TA, Gurevich I, Sakr R, Patil S, Stempel M, Morrow M. Occult malignancy in patients undergoing contralateral prophylactic mastectomy. *Ann Surg* 2011;254:2–7.

24. Foulkes WD, Reis–Filho JS, Narod SA. Tumor size and survival in breast cancer—a reappraisal. *Nat Rev Clin Oncol* 2010;7:348–53.

25. Dent R, Valentini A, Hanna W, *et al.* Young at local recurrence predicts breast cancer mortality. *Clin Oncol* 2014;:[in press].

26. Woods WG, Tuchman M. Neuroblastoma: the case for screening infants in North America. *Pediatrics* 1987;79:869–73.

27. Woods WG, Gao RN, Shuster JJ, *et al.* Screening of infants and mortality due to neuroblastoma. *N Engl J Med* 2002;346:1041–6.

28. Suita S, Zaizen Y, Sera Y, *et al.* Mass screening for neuroblastoma: quo vadis? A 9-year experience from the Pediatric Oncology Study Group of the Kyushu area in Japan. *J Pediatr Surg* 1996;31:555–8.

29. Narod SA. Age of diagnosis, tumor size, and survival after breast cancer: implications for mammographic screening. *Breast Cancer Res Treat* 2011;128:259–66.

30. Houvenaeghel G, Goncalves A, Classe JM, *et al.* Characteristics and clinical outcome of T1 breast cancer: a multicenter retrospective cohort study. *Ann Oncol* 2014;25:623–8.

Correspondence to: Steven A. Narod, Women's College Research Institute, Women's College Hospital, 790 Bay Street, Toronto, Ontario M5G 1N8.
E-mail: steven.narod@wchospital.ca

* Women's College Research Institute, Women's College Hospital, Toronto, ON.