

A Critical Appraisal of the Canadian National Breast Cancer Screening Study¹

BREAST cancer is the leading cause of death due to cancer in most countries of the Western world, and therapy has to date had little impact on mortality from the disease. The development of mammography offered the prospect of detecting breast cancer at an earlier stage than had previously been possible; hence, this suggested that the systematic use of mammography might lead to detection of a substantial proportion of breast cancers when they are still curable and to a reduction in mortality.

Several clinical trials have been conducted to test this hypothesis, and, while their results vary somewhat, they do in general indicate a reduction in breast cancer mortality as a result of screening with mammography. The magnitude of this reduction in women aged 50 years and older is about 30%; we are aware of no randomized controlled trial performed to date in younger women that showed a statistically significant reduction in mortality, except the Health Insurance Plan of New York. However, a recent meta-analysis of trials conducted in Sweden shows a 13% reduction in mortality for younger women (not statistically significant) (1).

A 30% reduction in mortality for a disease as common as breast cancer is of

enormous importance to the public health. It is not our intention here to review the results of these trials or of nonexperimental studies; this has been done at length in other recent publications (2,3). Rather, we will confine our attention to the recently published results of the National Breast Screening Study (NBSS) conducted in Canada (4,5) and consider whether the findings of that study should lead us to reconsider the view that mammographic screening reduces mortality from breast cancer. We do so using systematically the methods of critical appraisal (6) in which each of the major components of the trial is examined and its potential influence on the results considered.

GENERAL DESIGN OF THE NBSS

The NBSS was a randomized trial designed to answer two questions about screening for breast cancer: (a) Can screening with periodic physical examination and mammography reduce mortality from breast cancer in women aged 40–49 years? The comparison group for this section of the trial included women given a physical examination and taught breast self-examination at entry to the trial but not reexamined subsequently. (b) Can periodic screening with mammography and physical examination, compared with periodic physical examination alone, reduce mortality in women aged 50–59 years? The NBSS was, therefore, in effect two trials, each conducted in a specific age group. We consider these trials together but differentiate between them when necessary.

CRITICAL APPRAISAL

Selection of Subjects

Participants in the study were all volunteers who responded to one of the several methods of recruitment used in the trial. Thus, it cannot be assumed that the participants were representative of Canadian women, and they did differ from Canadian women in general in that the participants were more educated, had fewer children, and had a lower prevalence of cigarette smoking. These differences do not affect the validity of the internal comparisons made in the analysis of the trial; they may, however, influence the risk for breast

cancer and for death due to breast cancer in the trial. Women with symptoms of cancer were included, a feature of the trial that may have increased cancer incidence but that does not affect the validity of the group comparisons that are the subject of the analysis, provided there is no bias in the allocation of subjects (see Randomization).

Randomization

Participants were randomly assigned by the local coordinator from pre-printed lists supplied by the coordinating center. The allocations for each center were, therefore, available to the coordinator of each unit before they were used. This is not a method generally used in multicenter trials; reliance on a central source of allocation contacted by telephone is more usual. The shortcoming of the method of allocation used in the NBSS is that the allocation of individual subjects was known to the unit coordinator before consent was given. This foreknowledge offers opportunities for influence in allocation.

Miller et al (4,5) assert that no evidence of bias in allocation was found by the investigators. The evidence referred to includes the absence of erasures or omissions from the randomization lists and the similarity of the randomized groups in terms of risk factors for breast cancer. These findings do not alter the fact that it is impossible, when inspecting a list of names, to determine the order in which the names were entered. Whether and to what extent the randomization process followed the intentions of the investigators probably cannot definitively be determined this long after the event.

Evidence of imbalance in allocation is, however, present for the first (prevalence) screening examination that took place immediately after allocation. In women aged 40–49 years, 24 had breast cancer with poor prognosis, with four or more lymph nodes involved when they entered the trial. Nineteen of the 24 were allocated to the screening arm of the trial, and in all but two of them, breast cancer was detected at physical examination (4). Because physical examination was performed *before* randomization, the presence of these 17 tumors detected at physical examina-

Index terms: Breast neoplasms, 00.32 • Breast neoplasms, diagnosis, 00.111 • Cancer screening • Editorials

Radiology 1993; 189:661–663

¹ From the Division of Epidemiology and Statistics, Ontario Cancer Institute, 500 Sherbourne St, Toronto, Ontario, Canada M4T 1K9 (N.F.B., D.T., G.L.); the Department of Radiology, Mount Sinai Hospital, Toronto, Ontario, Canada (R.A.J.); Reichman Research Institute, Sunnybrook Hospital Health Sciences Centre, Toronto, Ontario, Canada (M.J.Y.); and the Department of Radiology, McMaster University, Hamilton, Ontario, Canada (C.J.Z.). Received February 24, 1993; revision requested March 23; final revision received July 6; accepted August 13. N.F.B. is the recipient of a National Health Scientist Award, Health and Welfare, Canada. **Address reprint requests to N.F.B.**

² Current address: Department of Diagnostic Radiology, Henry Ford Hospital, Detroit, Mich. • RSNA, 1993

tion was known at the time of randomization. We cannot say whether these findings influenced allocation, but only five women with palpable tumors were included in the control group at the prevalence screening. This imbalance could have arisen from bias in the allocation process or from chance.

The probability of an imbalance of at least this magnitude arising by chance can be calculated as a binomial (7) and is 3.3 in 1,000. Chance is, therefore, a possible but not likely explanation. Whatever the cause of this baseline imbalance in risk for death from breast cancer, the imbalance is likely to bias the estimate of relative risk for mortality from breast cancer in women aged 40–49 years. The direction of the bias will be to increase mortality in the mammography arm relative to the control arm.

Intervention

The interventions used in the NBSS were physical examination and mammography.

Quality.—The quality of mammography in the NBSS has already attracted considerable attention (8,9), and we do not intend to repeat what has earlier been stated. The evidence, however, suggests that the quality of mammographic images obtained in the NBSS did not represent state of the art even for the time. At most centers, the NBSS had to rely on the mammography equipment in place at inception of the study. The rapid pace of developments in mammography technology meant that a machine that was 5 years old in 1980 was markedly inferior to one manufactured in 1980. Many of the machines in use in the NBSS required long exposure times and used large-focal-spot x-ray tubes, and many centers lacked dedicated film processing. During the years in which the NBSS was carried out, major improvements were made in mammography technology, including the introduction and progressive refinement of antiscatter grids, improved automatic exposure control, and improved compression devices. Further, obtaining images in the mediolateral oblique view, which provides visualization of more of the breast than other views, was accepted as standard technique. In most cases, there was delay in incorporating these developments in NBSS centers.

In contrast, extensive attention was paid to the training of those persons who performed physical examination of the breast in the NBSS (10).

It is, nevertheless, to the credit of the investigators that they examined film quality and instituted a process of film review. How the method used by the investigators for the classification of film quality is related to the detectability of breast cancer, however, is not clear.

Evidence suggesting progressive improvement in image quality is available from the Center for Devices and Radiological Health, which monitors, in its Nationwide Evaluation of X-Ray Trends (NEXT) program (11), both radiation exposure and image quality in mammography and has shown progressive improvements from 1985 to 1992. To the extent that image quality has improved and thereby improved breast cancer detection, the NBSS results have limited application for modern practice.

Compliance.—Compliance is the extent to which the interventions were delivered as planned by the investigators. High levels of compliance are described: 86% of women aged 40–49 years and 87% of those aged 50–59 years underwent all five screening examinations. Compliance with physical examination in women aged 50–59 years was also high: 85% underwent all examinations. The effect of reduced compliance is to diminish the power of the trial by reducing the difference between groups in their exposure to the interventions being compared. No allowance for less than full compliance appears to have been made in the sample size calculations.

Contamination.—Contamination is the extent to which the control or comparison groups received the same interventions as the screened group. Of women aged 40–49 years, 26% of the control group underwent mammography at least once during the study period; of those aged 50–59 years, 17% of those allocated to undergo physical examination only also underwent mammography at least once during the study. The effect of contamination is to reduce the difference between groups in their exposure to the interventions being compared. This, in addition to the effects of noncompliance noted above, diminishes the power of the trial to detect differences in the interventions compared. No allowance for contamination appears to have been made in the sample size calculations.

Cointerventions.—Cointerventions are those treatments, apart from the interventions that are the subject of trial, that can influence the trial outcomes, in this case mortality due to breast cancer. In this context, treatment of detected breast cancer is the principal cointervention to be considered, and the investigators describe a review of treatment (not yet published) that failed to show any treatment-related differences that were associated with mortality from breast cancer.

Follow-up and Enumeration of Events

Complete follow-up is essential to ensure that all relevant events, in this case deaths due to breast cancer, have been identified. Follow-up mechanisms

used in the NBSS included mailed questionnaires and follow-up by physicians in those participants known to have developed breast cancer. Linkage to a national data bank was also used to identify subjects who had died of breast cancer. The extent to which these follow-up procedures were successful in establishing the health status of all NBSS participants at the time of analysis is not stated. Further, some of the mechanisms used for follow-up were complete only to 1988, and others to 1990. It is, therefore, not possible to determine from the published information exactly what period of follow-up was achieved. Nor is it possible, from the description given of a mean follow-up period of 8½ years, to determine if follow-up ended at the time of the analysis or on the date of the most recent positive verification of the health status of the subjects.

Deaths due to breast cancer or in women known to have breast cancer were all verified by independent review without knowledge of group assignment in the trial.

Statistical Considerations and Study Power

Results of both trials that made up the NBSS were negative in that the investigators failed to find statistically significant differences between the compared groups. The results are expressed as relative risk for dying of breast cancer. The relative risk is the ratio of the risk of dying in the group in each trial who underwent mammography divided by the risk of dying in the comparison group. These estimates were 1.36 for the trial in women aged 40–49 years and 0.97 for the trial in women aged 50–59 years. Neither of these relative risks differs significantly from 1. (A relative risk of 1 would indicate that the death rates in the two groups were the same.)

As with all negative trial results, we must consider how likely it is that the investigators could have detected the differences postulated at the time of the trial design and whether, because of an inadequate number of subjects or fewer events than anticipated, the trial results might be falsely negative.

The ability to detect the differences in mortality anticipated at the time of their formulation can be determined by reference to the confidence intervals around the point estimates of relative risk. These 95% confidence intervals were 0.84–2.21 for the women aged 40–49 years and 0.62–1.52 for those aged 50–59 years. The reductions in breast cancer mortality that were postulated when the trials were designed and their sample sizes determined were 40% with use of mammography plus breast physical examination in women aged 40–49 years and also 40% with use of mam-

mography plus physical examination of the breast, compared with physical examination alone, in women aged 50–59 years.

The NBSS was designed to have an 80% chance of detecting these large differences in mortality; however, substantially fewer deaths occurred than were predicted. For example, it was predicted that in women aged 40–49 years, there would be approximately 74 deaths over 7 years in the control group (12), whereas only 28 deaths had actually occurred at the time the results were published.

Had the anticipated reduction in death rates occurred, the relative risks would have been 0.6 in women aged 40–49 years and 0.6 in women aged 50–59 years. Both of these estimates are excluded by the 95% confidence intervals, and at first glance it appears that we can be confident that the NBSS result excludes benefits from use of mammography of the magnitude postulated. Whether this first impression is justified by the data is considered in the following section.

CONCLUSIONS

The NBSS is one of the largest trials yet performed to evaluate the efficacy of a screening test in reducing the risk for death due to a major health problem in Western society. Taken at face value, the results of the NBSS argue for abandoning mammographic screening as a population-based means of controlling death rates from breast cancer. We believe such a conclusion to be unjustified and unsupported by the findings of the NBSS.

The NBSS results in women aged 50–59 years, the age group for which there is strongest evidence of mortality reduction as a result of use of mammography to screen for cancer, are compatible with substantial reduction in mortality due to breast cancer attributable to screening with mammography. The 95% confidence intervals for this age group indicate that mortality reduction of up to 38% from the use of mammography are consistent with the NBSS data, as is a substantial increase in mortality.

For women aged 40–49 years, the NBSS results appear to exclude the anticipated 40% in mortality reduction. The 95% confidence interval, however, is compatible with a reduction in mortality as large as 16%. This result is likely biased by the large imbalance in women with breast cancer with poor prognosis at baseline. The confidence intervals around this biased estimate are, therefore, not a reliable guide to the magnitude of benefit that can be excluded in light of the NBSS findings. The influence of this baseline imbalance will diminish as the number of women in the trial who die of breast cancer increases, and extensive further follow-up will be required to determine whether the present estimates of relative risk change. Uncertainty about the magnitude of benefit from screening in the NBSS may remain, however, because of lack of information about the continued use of mammography in former members of the trial.

Until such time as further follow-up and analysis reduce the uncertainty surrounding the magnitude of the effects on mortality seen in the NBSS, the results of these trials should not be used to change the prevailing scientific view of the potential benefits of screening with mammography. ■

References

1. Nystrom L, Rutqvist LE, Wall S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993; 341:973–978.
2. Hurley SF, Kaldor JM. The benefits and risks of mammographic screening for breast cancer. *Epidemiol Rev* 1992; 14:101–130.
3. Day N. Screening for breast cancer. *Br Med Bull* 1991; 47:400–415.
4. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study. I. Breast cancer detection and death rates among women aged 40 to 49 years. *Can Med Assoc J* 1992; 147:1459–1476.
5. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study. II. Breast cancer detection and death rates among women aged 40 to 49 years. *Can Med Assoc J* 1992; 147:1477–1488.
6. Sackett DL, Haynes RB, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. Boston, Mass: Little, Brown, 1985.
7. Snedecor GW, Cochran WG. *Statistical methods*. 6th ed. Ames, Ia: Iowa State University Press, 1967.
8. Kopans DB. The Canadian screening program: a different perspective. *AJR* 1990; 155:748–749.
9. Day NE, Duffy SW. Breast screening in women under 50 (letter). *Lancet* 1991; 338:113.
10. Baines CJ, Miller AB, Bassett AA, et al. Physical examination: evaluation of its role as a single screening modality in the Canadian National Breast Screening Study. *Cancer* 1989; 63:160–166.
11. Conway BJ, Suleiman OH, Rueter FF, Antonsen RG, Slayton RI, McCrohan JL. Does credentialling make a difference in mammography? (abstr). *Radiology* 1992; 185(P):393.
12. Miller AB, Howe GR, Wall C. Protocol for a Canadian randomized controlled trial of screening for breast cancer in women. *Clin Invest Med* 1981; 4:233–246.