The review of randomization in the Canadian National Breast Screening Study

Is the debate over?

Norman F. Boyd, MD

Editorial

Éditorial

Dr. Boyd is with the Ontario Cancer Institute, Princess Margaret Hospital, Toronto, Ont.

Can Med Assoc J 1997;156(2): 207-9.

Abstract

THE RANDOMIZATION PROCEDURE in the Canadian National Breast Screening Study (NBSS) is assessed in this issue (see pages 193 to 199) by Drs. John C. Bailar III and Brian MacMahon. They conclude that although there was ample opportunity for the randomization process to be subverted, no evidence of subversion was found. This is unlikely to allay all concerns about randomization, because there are still puzzling differences between the arms of the NBSS in a number of baseline variables. For example, the existence of prior health claims for breast cancer for women who entered the NBSS in Manitoba has raised the possibility that subversion occurred. Although the question may never be resolved, one lesson is clear: randomization in clinical trials should be managed in a manner that makes subversion impossible. As for the clinical implications of the NBSS for women in their 40s, physicians may now look to the results of randomized trials that have been published more recently. A meta-analysis of these results suggests that screening mammography reduces deaths from breast cancer among women in their 40s, but continued follow-up over the next few years will be needed to settle the debate.

Résumé

DANS LE PRÉSENT NUMÉRO (voir pages 193 à 199) les D[rs] John C. Bailar III et Brian MacMahon évaluent la méthode de randomisation utilisée dans le cadre de l'étude nationale sur le dépistage du cancer du sein au Canada. Ils concluent que, malgré les nombreuses possibilités de corruption des méthodes de randomisation, on n'a trouvé aucune preuve de corruption. Cette conclusion a peu de chances de dissiper toutes les préoccupations suscitées par la randomisation, car il reste toujours des différences intriguantes entre les divers volets de l'étude nationale pour plusieurs des variables de référence. Par exemple, l'existence de réclamations antérieures liées au cancer du sein chez des femmes qui ont participé à l'étude au Manitoba a soulevé la possibilité de corruption. Même s'il se peut que l'on ne règle jamais la question, une leçon à en tirer est claire : il faut gérer la randomisation dans le contexte d'études cliniques de façon à rendre la corruption impossible. Quant aux répercussions cliniques de l'étude nationale sur les femmes dans la quarantaine, les médecins peuvent maintenant consulter les résultats d'études randomisées publiés plus récemment. Une méta-analyse de ces résultats indique que les tests de dépistage de la mammographie réduisent les décès causés par le cancer du sein chez les femmes dans la quarantaine, mais il faudra poursuivre la recherche au cours des prochaines années pour trancher la question.

Let me start by stating that I am not a disinterested observer of the Canadian National Breast Screening Study (NBSS). From 1976 to 1978 I worked under the direction of Dr. Anthony Miller in the Epidemiology Unit of the National Cancer Institute of Canada, from which the NBSS was designed and coordinated, and carried out pilot studies in preparation for the NBSS.[1,2] I have used, and continue to use, NBSS data for epidemiological purposes with the collaboration of Miller,[3–6] and, with colleagues, I have published a critical appraisal of the NBSS.[7] However, I do not have any vested interest in the effectiveness of screening mammography.

When the NBSS was started, there was no evidence available from random-

ized studies of the role of screening for breast cancer in women under the age of 50, nor was there evidence of the magnitude of the separate contributions of physical examination and mammography to the reduction in the rate of death from breast cancer among women over 50.

The NBSS was designed to fill these gaps in our knowledge. A key aspect of its design was that the trial be randomized. Because the main purpose of the study was to compare the numbers of women who died of breast cancer in the mammography arm and the control arm (usual care), it was essential that the two groups be comprised of women whose risk of death from breast cancer, without screening, was the same. Randomization, properly carried out, would ensure such similarity and allow unbiased assessment of the effectiveness of the screening intervention.

In this issue (see pages 193 to 199) Drs. John C. Bailar III and Brian MacMahon assess whether randomization was properly carried out in the NBSS. Evidence of subversion would call into question the credibility of the results of this expensive and time-consuming study.

The possibility that randomization in the NBSS may have been subverted was discussed and rejected by Miller and associates in the 1992 article reporting the 7-year follow-up results of the NBSS.[8] In their discussion, they noted that the method of randomization used (described in Bailar and MacMahon's report) meant that the coordinator responsible for randomization at each study centre knew, or had access to, the randomization schedule before subjects were actually assigned to one of the two groups. Thus, after each subject had been examined by a nurse (or a physician, in Quebec), but before randomization took place, the coordinator knew the group to which the next subject would be allocated. In principle, subversion of the sequence of events prescribed by the investigators could thus have taken place. Miller and associates noted the similar rates of referral for diagnostic evaluation and the similar distribution of risk factors, and other evidence, and concluded that there was no evidence of subversion.

Questions about randomization in the NBSS,[7,9] however, have subsequently been raised because of the study's findings that showed an imbalance in the number of women with advanced breast cancer. In particular, the mammography arm had an excess of patients with involved axillary lymph nodes whose breast cancer had been detected by physical examination at baseline (i.e., before randomization). The signs of breast cancer detected by physical examination were thus known before randomization: in the group 40–49 years old, 17 such women were allocated to the mammography arm and 5 to the control arm (Table 1). We obviously do not know whether the physical findings influenced allocation, but given how randomization was carried out, they could have. The imbalance may also have arisen by chance. However, from the results of the binomial probability calculation in Table 1, this is not likely.

The issues Bailar and MacMahon address are deliberately limited in scope. They examined alterations of subjects' names in the "allocation books" as well as events that occurred during a period when administrative problems were reported at one of the study centres. They further limit their attention to the subjects 40–49 years of age, the group in which there was an apparent excess of deaths from breast cancer in the mammography arm. They present convincing evidence that there was no association between alterations, or administrative problems, and the rate of death from breast cancer.

These findings, however, are unlikely to quell completely the concern about the randomization process in the NBSS. The absence of name alterations had previously been cited by the NBSS investigators as evidence that randomization had not been subverted.[8,11] We now know that names were altered and that there were more alterations in the mammography arm. Although 78% of these changes could be accounted for in some way (e.g., clerical errors), the remaining 22% (representing 101 names) could not. We know nothing of who these women were or why their names appeared once in the allocation books, were replaced by another name and never appeared again in the NBSS. We also now know that a coordinator at one of the NBSS centres was suspected of assigning her friends to the mammography arm and that the suspicion was strong enough to remove her from her position. We do not know the method of subversion thought to have been used by

Table 1: Frequency of selected events in the Canadian National Breast Screening Study, by study arm

		Study arm; no. of subjects		
Event	Age	Mammography	Usual care	*p* value
Prior health claim for breast cancer[10]	40–59	8	1	0.05†
Alteration of name in allocation book*	40–49	95	65	0.01
Advanced breast cancer detected at baseline by physical examination[7–9]	40–49	17	5	0.003
Death from breast cancer in 7-year follow-up period[8]	40–49	38	28	0.27†

*Data from the review by Bailar and MacMahon (see pages 193 to 199).
†Calculated by the χ^2 test.

this coordinator, nor do we know whether an examination of name alterations revealed what she was alleged to have done. No coordinators were interviewed. Although it is unlikely, as Bailer and MacMahon suggest, that any would admit wrongdoing, if such admissions had been made they would have provided powerful evidence.

Apparently, great deviousness would not have been required to achieve a particular allocation. Simple, risk-free methods appeared to exist if a subject wished, with the cooperation of the coordinator at the centre, to be allocated to one of the two study arms. Suppose a subject wished to be allocated to the mammography arm? She would have had a 50/50 chance of being assigned to that group in any event. If the next allocation was to the control arm instead, the subject's name could have been entered onto the line with the next mammography allocation, leaving a gap in the allocation book, or she could have been advised to wait until the line for the desired arm was the next to be filled. In either case, it is unlikely that much time would have elapsed before a mammography allocation came up or a gap on the list was filled. Fifteen NBSS centres randomly assigned 90 000 women over 5 years or less: they must have been busy places.

A recent article adds to the concerns about randomization in the NBSS trial. Cohen and associates[10] examined previous health claims submitted for subjects who entered the NBSS centre in Manitoba. They found no statistically significant differences between the mammography and control arms for a large number of prior conditions, including several types of investigation for breast disease. However, 9 prior claims for breast cancer were found among the NBSS participants in Manitoba. Of these individuals 8 were assigned to the mammography arm and 1 to the control arm ($p = 0.05$). Of the women 40–49 years old, all 4 who had prior claims for breast cancer were assigned to the mammography arm ($p = 0.12$).

Table 1 summarizes the published evidence from the NBSS that is at least consistent with the possibility that randomization may have been subverted.

Differences exist between the mammography and control arms in the number of prior claims for breast cancer, of name alterations in the allocation books and of women with advanced breast cancer detected at baseline by physical examination. Explanations for each of these findings have been offered. For the prior health claims for breast cancer, no other evidence could be found to support the diagnosis.[10] For the difference in the number of name alterations, subjects in the mammography arm were seen more often by the centre staff and so there were more opportunities for corrections to be made. For the difference in the number of women with advanced cancer, Miller and associates[8] pointed out that mammography contributed to the "greater efficiency in cancer detection in the mammography group." Because so much time has passed since the events in question occurred, it may not be possible to determine definitively whether these various explanations are correct, or whether these differences are due to the single explanation that randomization was subverted. Interested readers should consult the original sources to judge for themselves.

One lesson to be learned from the NBSS is clear: randomization in clinical trials should be managed in a manner that makes tampering of any sort impossible. This could be achieved in multicentre trials like the NBSS if randomization were performed by telephone contact with a site remote from the centres where subjects would be enrolled. In this way, people responsible for enrollment would have no advance knowledge of allocation.

What are the implications of the NBSS findings for breast cancer screening? The state of knowledge is now very different from what it was in the late 1970s, when the NBSS was conceived. Reports of several randomized trials of screening with mammography among women under 50 have now been published, and a meta-analysis of their findings suggests that mammography is effective in reducing the rate of death from breast cancer in this age group.[12] Continued follow-up in these trials over the next few years should settle the debate.

References

1. Boyd NF, Sutherland HJ, Fish EB, Hiraki GY, Lickley HLA, Maurer VE. A prospective evaluation of physical examination of the breast. *Am J Surg* 1981;142(3):331-4.
2. Boyd NF, Wolfson C, Moskowitz M, Carlile T, Petitclerc M, Ferri HA, et al. Observer variation in the interpretation of xeromammograms. *J Natl Cancer Inst* 1982;68(3):357-63.
3. Boyd NF, Wolfson C, Moskowitz M, Carlile T, Petitclerc C, Ferri H, et al. Observer variation in the classification of mammographic parenchymal pattern. *J Chronic Dis* 1986;39(6):465-72.
4. Boyd NF, Jensen H, Cooke G, Lee Han HW. The relationship of mammographic and histological risk factors for breast cancer. *J Natl Cancer Inst* 1992;84:1170-9.
5. Lee Han H, Jensen H, Cooke G, Boyd NF. A comparison of the epidemiology of atypical hyperplasia of the breast epithelium and mammographic parenchymal densities. *Eur J Cancer Prev* 1993;2(suppl 3):33-40.
6. Boyd NF, Byng J, Yaffe M, Jong R, Fishell E, Little L, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *J Natl Cancer Inst* 1995; 87:670-5.
7. Boyd NF, Jong RA, Yaffe MJ, Tritchler D, Lockwood G, Zylak CJ. A critical appraisal of the Canadian National Breast Screening Study. *Radiology* 1993; 189:661-3.
8. 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. *Can Med Assoc J* 1992;147:1459-76.
9. Tarone RE. The excess of patients with advanced breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. *Cancer* 1994;75(4):997-1003.
10. Cohen MM, Kaufert PA, MacWilliam L, Tate RB. Using an alternative data source to examine randomization in the Canadian National Breast Screening Study. *J Clin Epidemiol* 1996;49(9):1039-44.
11. Baines CJ. The Canadian National Breast Screening Study: a perspective on criticisms. *Ann Intern Med* 1994;120(4):326-34.
12. Smart CR, Hendrick RE, Rutledge JH III, Smith RA. Benefit of mammography screening in women age 40–49: current evidence from randomized controlled trials. *Cancer* 1995;75:1619-26.

Reprint requests to: *Dr. Norman F. Boyd, Ontario Cancer Institute, Princess Margaret Hospital, 610 University Ave., Toronto ON M5G 2M9*