Most Breast Cancer Screening Trials Have a Flawed Design

Nishant Gurnani · Anurag Srivastava

Received: 17 January 2011 / Accepted: 27 September 2011 / Published online: 25 October 2011
© Association of Surgeons of India 2011

Abstract In the present article, we discuss that why most breast cancer screening trials have a flawed origin. We suggest some solutions to correct these flaws so that more valid and reliable screening trials can be conducted in the future.

Keywords Breast cancer · Screening trials · Bias · Quality of life · Mammography

Introduction

Breast cancer is the leading cause of cancer-related death among women. Approximately 1 in 8.2 women will receive a diagnosis of breast cancer during her lifetime, and 1 in 30 will die of the disease [1]. Breast cancer incidence increases with age [1], and although significant progress has been made in identifying risk factors and genetic markers, more than 50% of the cases occur in women without known major predictors [2–5]. Hence, there is a need to detect breast cancer early by screening methods to improve the outcome of breast cancer. Here, we review why most of the screening trials are flawed in design and what measures should be taken while conducting the future trials.

Review of Breast Cancer Screening Trials: The Situation So Far

There have been various studies regarding the efficacy and effectiveness of breast cancer screening trials. Throughout

Europe—including Denmark [6], Italy [7], Norway [8], Sweden [8], and the United Kingdom [9]—the initiation of screening mammography has been associated with increased breast cancer incidence among the women of screening age.

There has been a review article [10] of Canadian National Breast Screening Study in which critics have attacked the study's design, randomization, execution of mammography, follow-up procedures, contamination of controls, and analysis. The absence of benefit observed in the mammographically screened women who were 50–59 years old has been used to support the criticism of mammography. According to this review article randomization was flawed and more symptomatic women were assigned to mammography arm of trial.

In many studies [11], mammography has not been found to be an effective screening tool in younger women. This is because of the greater difficulty in distinguishing the normal tissue from the abnormal one on the mammographic images, and because of the greater growth rate in younger women.

There are certain flaws that need to be rectified while conducting the future screening trials.

The “Flaws” in Screening Trials

The most robust evidence in favor of the beneficial effects of population-based or hospital-based screening is derived from the meta-analysis of large randomized trials.

However, randomized screening trials are bothersome. They require a very long follow up period to demonstrate a benefit of screening in terms of reduction in morbidity and mortality. Moreover, the trials need to be large in size to be able to answer the questions with high statistical power. A sufficient number of persons in the study arm have to be screened, but if randomization has been carried out before

N. Gurnani · A. Srivastava (✉)
Department of Surgery, All India Institute of Medical Sciences,
New Delhi, India
e-mail: dr.anuragsrivastava@gmail.com

consent (which is possible, depending on the local country-specific regulations), attendance for screening may be low in practice [12]. If randomization is carried out after informed consent, it may affect the attitude of persons randomized to the control arm to move to the mammographic arm. The long time frame of the trials, in general, has the disadvantage that newer techniques may have evolved in the meantime, urging investigators to change the screening protocol and therefore possibly weakening the final results of the trial [13].

There are eight large RCTs published, out of which four have been conducted in the Sweden [12]. An overview of these Swedish trials was published in the *Lancet* in 2002.

Most of the trials suffer from the following methodological drawbacks:

1. Randomized trials are essentially prospective cohort studies in which the intervention has been allocated by the investigator employing the process of randomization. The basic tenet of any cohort design is that at the time of recruitment of subjects in the study, the subjects should be free from the disease or the outcome of interest. If the outcome of interest is death, the subjects should be alive on the inception of cohort. All the potential subjects (after the informed consent), should be screened for the disease or the outcome of interest. The screening test should have very high sensitivity in picking up the disease, ideally the “gold standard” available at the time of the study. The same gold standard test should be applied to all the subjects—test subjects as well as the controls. If one test is applied to intervention or the test arm and another to the controls, a differential misclassification type of measurement bias will be introduced [14].

In almost all the screening trials of breast cancer, the mammography arm patients were screened by a mammogram at the inception of the cohort and later at varied intervals. The subjects in the control arm only received a physical examination, and in some studies education about performing BSE. The mammogram was not offered to women in the control arm. At the time of the above trials, mammogram was considered the most sensitive test of picking up breast cancer. Since control arm did not receive it, some cancers would have been missed and women with these cancers included in the study. These cancers were then allowed to grow and manifest when symptomatic. The outcome of therapy of symptomatic tumors is known to be worse because these are of larger size and often have spread to the axillary nodes. Today MRI scan is considered the most sensitive test for picking up breast cancer and any screening trial starting today should have a baseline screening of all subjects (cases as well as ladies in the control arm) by a MRI scan.

2. Prevalent versus incident cases: Another principle of a cohort study is to include only incident cases, that is, new cases arising within in the study period. Since a sensitive test was not applied to women in the control group, both prevalent as well as incident cases were included in the controls as opposed to only incident cases in the intervention arm. Thus, control arm will include more cancer cases, and poor outcome in this arm may simply be due to more cancer cases in the controls.
3. Wrong outcome: Screening is essentially a diagnostic test which is meant to pick up disease in the asymptomatic individuals. Screening on its own will not change the outcome. The outcome depends on the prognostic factors present at the time of diagnosis. The screening can alter these factors by detecting the disease early when two most important prognostic factors, that is, tumor size and nodal spread are still favorable. Hence, the appropriate outcome of a screening study should be the number or proportion of cases detected with sub-centimeter tumor, number of grade I tumors, number of cases without lymphatic spread (with sentinel node negative tumors), number of estrogen-receptor-positive tumors, and number of women detected with a “triple-receptor-negative (negative for estrogen receptor, progesterone receptor, and HER-2 receptors) tumors.
4. The outcome of a patient with breast cancer also depends on the extent and appropriateness of surgery for breast and axilla; timing and dosage of radiotherapy; type, dosage, and duration of chemotherapy; and hormone therapy. Since these therapies are undergoing rapid changes, the outcome of patients detected through screening is also going to change. We are already witnessing improvement in the mortality of breast cancer in the UK and in the USA. Hence, we should continue screening even if today there is little or no survival advantage of screening in a particular age group (viz. age group 40–49 years).
5. Most oncologists are in the habit of measuring death as the outcome of breast cancer [15]. The precise cancer-specific or cancer-related mortality is impossible to measure, unless we perform autopsies in all the dead cases. Even on autopsy, a small focus of cancer may be missed. Even if a histologically confirmed metastasis is revealed on autopsy one can never be 100% certain whether it was a “death due to cancer” or “death with cancer”. The overall survival in a cancer study is not synonymous with effectiveness of therapy. For a cancer with equal lethality in all ages, the mortality would be higher in the older population because of concomitant illnesses. Some authors try to correct for this by determining the cause of death and analyzing the “cause specific mortality”. Unfortunately, the determination of the cause of death is not always accurate. In order to avoid the need of distinguishing death caused by breast

cancer as opposed to death because of other causes, Saphner and coworkers [16], chose to calculate the annual hazard of recurrence as opposed to survival in the group of women entered in to “Eastern Cooperative Oncology Group” studies on breast cancer.

According to Baum [17] the reduction in breast cancer mortality owing to screening is much less than it has been credited for. Furthermore, the harms from false-positive results and the over-diagnosis of indolent disease, which includes the detection of a cancer that is not destined to present clinically in that patient’s lifetime, are now perceived as much greater than ever anticipated.

6. The terms lead time bias and length time bias are theoretical constructs that may not be applicable to non-lethal cancers and precancers detected today with modern screening tools. Modern breast cancers that are detected early are picked up in initial microinvasive or preinvasive stages where they do not cause mortality. In the Service screening of National Health Services of Great Britain, which started in 1995, more than 90% women are alive and well at the end of 10 years follow-up. Most of these women will probably live for another 10 years, without dying. In such women, there is no question of a lead time or length time bias. The calculation of the lead time and length time biases uses “date of death” as the end point of interest [18].

What to Do: Suggested Solutions to Correct the Flaws

1. In a trial, the most sensitive technique available should be offered to both the control and experimental group at the inception of study in order to detect the “prevalent cancer cases”, which should be (appropriately treated and) excluded from the RCT.
2. We must measure the following parameters to evaluate the success of breast screening:
 - a. The “breast salvage years” gained (similar to person years concept followed in calculating incidence density, because with detection of early cancer, breast will be salvaged and remain salvaged till a recurrence necessitates a mastectomy) because of early detection of cancer. For example, if in a cohort of 1,000 ladies, 100 cancers are detected, 80 in early stage and 20 with advanced stage. Of the 80 in early operable cancer cases, 70 are treated by breast conservation therapy. Among the 20 advanced tumours treated with neo-adjuvant chemotherapy, 10 later undergo breast conservation therapy. So, the initial breast salvage rate would be 80/100 or 80%. If later 10 of these 80 ladies develop a loco-regional failure and then undergo a completion mastectomy for a recurrence at say year 3. Thus, the breast salvage rate would be computed as follows:

In first 3 years $80 \times 3 = 240$ person years

At year 3; 10 undergo mastectomy leaving 70 cases, so salvage for next 2 years $= 70 \times 2 = 140$ person years; adding these two values we get 380 person years of breast salvage rate for five years of follow up. If a screening programme detects more of early cancers which are suitable to breast conservation therapy, and these are predominantly node negative good prognosis tumours which remain recurrence free, the “breast salvage rate” achieved by this screening programme would be very high.
 - b. The “axilla salvage years” gained: Among women detected with early breast cancer with node negative axilla (on clinical palpation and axillary ultrasound scan) a sentinel lymph node biopsy is performed. In women with a histologically negative sentinel lymph node, the axilla is spared or salvaged from undergoing a full axillary nodal removal which adds to the cost of the procedure, extra work to the surgeon, risks of lymphoedema, shoulder pain, paresthesia, impaired body function and reduced quality of life.). Thus, we can compute “axilla salvage years” gained among screened women, which will remain salvaged till an axillary recurrence necessitates full axillary lymph node dissection. Its computation will follow the same method as described above for “breast salvage years”.
 - c. Number of women in whom cytotoxic chemotherapy is avoided: This should be considered an important societal gain. Women presenting with a sub-centimeter tumor (<1 cm diameter on final pathology), with grade I and lymph node-negative disease and those with Ductal carcinoma in situ, do not require systemic chemotherapy. Given the choice, all ladies want to avoid chemotherapy. All girls abhor loosing their beautiful hair. There is a significant cost saving also by avoiding chemotherapy.
 - d. Cost-effectiveness of therapeutic program: With the present cost escalation, we should be cognizant of the high cost of treating an advanced cancer as opposed to an early disease. The surgery for advanced cancer with a large lump with a bleeding ulcer cannot offer long-term control even with expensive anticancer drugs and radiotherapy. These women with locally advanced cancers require prolonged operation often with some plastic surgical flap repair followed by expensive and very toxic anticancer drugs and radiotherapy. The total treatment takes 6–9 months with a poor quality of life and huge economic burden,

and often debt to the family. After all this the poor patient develops local recurrence in 6–18 months, requiring more chemotherapy with higher cost, followed by emergence of systemic deposits in 2–3 years and death in 4–5 years. The quality of life usually remains very poor during this entire episode of cutting (surgery), poisoning (chemotherapy), and roasting (radiotherapy), which does not help the patient much. The family is posed with additional economic burden in bearing the cost of care and need of a home help and lost wages. In brief, the therapy of a case with advanced cancer is a societal waste. It neither helps the patient nor the society and taxes the exchequer fruitlessly.

3. Care must be taken to prevent contamination of control groups. If it does happen, as it will in real life, the number of women receiving breast screening tools in the control arm and the number of women refusing a screening tool in the test arm should be recorded, so that appropriate analysis could be carried out.
4. Most importantly, we must look at the following parameters of life and not death in cancer patients [15]:
 - a. Presence of local recurrence of tumor confirmed by histology.
 - b. Presence of metastasis to vital organs such as lung, liver, and brain confirmed by MR/CT/isotope scan/histology.
 - c. Health performance score using standard “European Oncology Research and Treatment group” guidelines.
 - d. The Quality of life years (QUALY) gained or the disability affected life years (DALY) decreased can be measured [19].

Conclusion and Recommendation

Most breast cancer screening trials have a flawed design or analysis. These flaws are due to improper screening methods being utilized for control and experimental groups. Moreover, the success or effectiveness of screening should be assessed by parameters of life, and not death. More stress should be given to improving the quality of life of the whole society (patient, the family, and the society she lives in) as a result of early cancer detection and costs saved rather than prevention of deaths.

Source of Funding Nil

Conflict of Interests None

Previous presentation or publication None

References

1. Cancer Facts and Figures (2001) American Cancer Society. Accessed at www.cancer.org/downloads/STT/F&F2001.pdf on 16 July 2002.
2. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C et al (1989) Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 81:1879–1886
3. Colditz GA, Willett WC, Hunter DJ, Stampfer MJ, Manson JE, Hennekens CH et al (1993) Family history, age, and risk of breast cancer. Prospective data from the Nurses’ Health Study. *JAMA* 270:338–343
4. Seidman H, Stellman SD, Mushinski MH (1982) A different perspective on breast cancer risk factors: some implications of the nonattributable risk. *CA Cancer J Clin* 32:301–313
5. Strax P (1987) Mass screening of asymptomatic women. In: Ariel IM, Cleary J (eds) *Breast cancer: diagnosis and treatment*. McGraw-Hill, New York, pp 145–151
6. Zahl P-H (2004) Overdiagnosis of breast cancer in Denmark [letter]. *Br J Cancer* 90(8):1686
7. Paci E, Warwick J, Falini P, Duffy SW (2004) Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? *J Med Screen* 11(1):23–27
8. Zahl P-H, Strand BH, Mæhlen J (2004) Breast cancer incidence in Norway and Sweden during introduction of nation-wide screening: prospective cohort study. *BMJ* 328(7445):921–924
9. Vainio H, Bianchini F, Heseltine E (2002) IARC handbooks of cancer prevention. Vol 7: breast cancer screening. International Agency for Research on Cancer, Lyon, p 147
10. Baines CJ (1994) The Canadian National Breast Screening Study: a perspective on criticisms. *Ann Intern Med* 120(4):326–334
11. Elwood JM, Cox B, Richardson AK (1993) The effectiveness of breast cancer screening by mammography in younger women. *Online J Curr Clin Trials* 1993 Feb 25; Doc No 32
12. Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE (2002) Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 359(9310):909–919
13. de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ (1995) Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. *J Natl Cancer Inst* 87(16):1217–1223
14. Rothman KJ, Greenland S. In *modern epidemiology*. Precision and validity of studies, 2nd edn. Lippincott Raven Publishers, Philadelphia, pp 126–127
15. Srivastava A, Sood A (2009) Oncologists should measure life, not death: a newer perspective of cancer statistics. *Indian J Cancer* 46:13–16
16. Saphner T, Tormey DC, Gray R (1996) Annual hazard rates of recurrence for breast cancer after primary therapy. *J Clin Oncol* 14:2738–2746
17. Baum M (2010) Should routine screening by mammography be replaced by a more selective service of risk assessment/risk management? *Women’s Health* 6(1):71–76
18. Baker SG, Kramer BS, Prorok PC (2002) Statistical issues in randomized trials of cancer screening. *BMC Med Res Methodol* 2:11. doi:10.1186/1471-2288-2-11
19. Murray C, Lopez A (1997) Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. *Lancet* 349(9064):1498–1504