

Commentary: False premises, false promises and false positives—the case against mammographic screening for breast cancer

Michael Baum

Before I start it is important that I establish my credibility as my somewhat eccentric career allows me an unusual insight into the problem. First of all I come from a family with a bad history of breast cancer, which in part fuelled my zeal for entering this sub-specialty. Next I am a surgeon who has been at the front line of breast cancer treatment for over 30 years and during the latter half of that experience have been looking after the ‘casualties’ of the screening programme in the UK. I know a lot about the mechanics of screening, having been given the task of setting up the programme in the South East of England in my role as Professor of Surgery at Kings College Hospital in 1987/88 and then serving on our national screening committee until 1997. Finally, I am well equipped to discuss the minutiae of randomized controlled trials (RCT) as founder of the first clinical trials centre in the UK in 1980 and as principle investigator of the largest collaborative group for breast cancer trials in Europe, until my retirement in June of this year.¹ I know a lot about the quality issues in running RCT and my group established many of these principles in the late 1970s.

Let us start therefore with a check-list of some of the essentials for running high quality RCT that are now enshrined in the CONSORT principles on the conduct and reporting of such trials.²

- The study population must be precisely defined for prospective entry and exclusion criteria.
- All patients must be recruited with individual informed consent.
- Randomization must be blinded in such a way that it is impossible for the investigator to predict what the next allocation is going to be.
- There should be *a priori* power calculations based on the primary end points in order to avoid type I and type II errors.
- The primary endpoints should be ‘event free’ survival, which include deaths from other causes before a breast cancer event as well as breast cancer specific events. This analysis should be triggered at a predetermined threshold on the advice of an independent data monitoring and safety committee (DMSC).
- The secondary endpoints should be all-cause mortality and cause-specific mortality again triggered by the advice of the DMSC.
- A priori* ‘stopping rules’ should be activated by the DMSC if there are unexpected adverse events or if the advantage of one arm over the other becomes apparent earlier than expected, making the appropriate allowances in the *P*-values for multiple interim analyses seen only by the DMSC.
- Analysis should be based on intention to treat only.

—Adverse events linked to the intervention should be formally collected prospectively in order to facilitate a harm/benefit analysis for clinical utility.

According to this check-list none of the clinical trials of mammographic screening meet all the criteria, although the Canadian studies come close. So what we are witnessing is a double standard in that clinical trials of screening are acceptable at a quality that would completely invalidate trials for the prevention or treatment of breast cancer. Yet those who to dare to criticize them are treated as pariahs.

Let me expand on the importance of all-cause mortality as an endpoint and the need to analyse by intention to treat (ITT).

No intervention dreamt up by man has only the power of good without the risk of inadvertent harm.³ The bland assumption that the process of diagnosis and treatment of screen-detected lesions in the population at large, some of which might never have expressed a malignant potential in the woman’s natural lifetime, is totally free of risk to life, is breathtaking in its arrogance. When the absolute benefit for saving life from breast cancer is estimated at about one in 1200 women over 14 years,⁴ during which time countless numbers of women would have had invasive procedures to exclude breast cancer or had been over-treated for borderline pathology, only one adverse event in 17000 woman’s years of screening, i.e. 1200×14 , would wipe out that gain. I accept that the power calculations for all-cause mortality would have suggested many more volunteers and huge additional costs to allow for this; to which I respond by saying either don’t embark upon it in the first place or ‘go for it’ as in the women’s health initiative (WHI) studies on hormone replacement therapy.⁵

Coming now to the importance of ITT analysis. You simply cannot assume that the ‘acceptors’ and the ‘refusers’ of screening have the same risk of dying of breast cancer. ‘Compliant’ acceptors tend to be more health aware and of a higher social class than ‘non-compliant’ refusers. These factors alone are known to be strongly associated with the outcome of treatment for breast cancer. Furthermore in pragmatic trials of population interventions the acceptability of the intervention is on trial as much as its outcome. You cancel the party if too many people refuse the invitation rather than saying what a good time they could have had if only they accepted!

I now wish to turn to some of the specifics in the paper by Freedman, Petiitti, and Robins.⁶ For a start I found many of their arguments difficult to follow in common with most of the reworking and defences of the old screening trials.⁷ Maybe this is because I am not very bright or maybe it is because of the contrast between the transparency of well-conducted RCT and the opacity of the trials of mammographic screening.

In their introduction they claim that one of the objectives of screening is to allow 'treatment that is less invasive'. Well that is a false claim for a start. In the audit of surgery for the National Health Service Breast Screening Programme (NHSBSP) it is reported that between 20 and 50% of cases of duct carcinoma *in situ* (DCIS) ('early breast cancer'?) end up having mastectomy⁸ and the net effect of NHSBSP on invasive surgical procedures on the breast in England and Wales has been an increase of just under 20% mastectomy rates for invasive breast cancer and over 400% mastectomy rates for DCIS since 1990 corrected for any changes in demography of the population.⁹

In defense of the Health Insurance Plan (HIP) study they claim that the two groups **should** be equal because 'women were assigned to study or control in alternation'. As this is an unblinded allocation this trial doesn't get past first base and should be excluded as a brave but ultimately futile exercise. It is also curious that the study with highest relative risk reduction in breast cancer deaths should be the one with the most primitive radiological techniques with only a 5% detection of DCIS compared with 20% in more recent studies. They also prove my point on the importance of absolute rather than relative risk reductions by trying to make a virtue of the fact that 24 fewer breast cancer deaths after screening more than 30 000 women for four rounds translates into a 62% relative risk reduction. Next there is the knotty problem of excluding patients with a prior diagnosis of breast cancer. This should have been an *a priori* exclusion criterion in a valid study. To claim that to do it properly would have been expensive is a pretty feeble excuse. Do it properly or not at all. No amount of *post hoc* massaging of the data can compensate for this fundamental error. It would be equivalent of including patients in a chemoprevention trial with a past history of breast cancer and waiting for a recurrence before excluding them from future analysis!

Coming now to the Two-County trial. Again the method of allocation to the study or control group would not pass muster these days but at least it was less open to abuse than the HIP study leaving the residual problem of demographic matching.

I am also not clear why the trial was closed after four or five rounds. Was this on the advice of the DMSC after a predetermined number of events? If not could this have been on a random high?

Unlike the HIP study, the quality of radiology was excellent, yet the Two-County trial group is for ever broadcasting the results in terms of case survival (ignoring lead time bias) or as a sub-group analysis of acceptors, ignoring ITT. Freedman *et al.* appear to make a virtue of this, completely ignoring one of the fundamental principles of the philosophy of clinical trials. The statement they make that 'lead time bias means that screening speeds up detection, so incidence is higher in the study group at the beginning of follow up' suggests a misunderstanding of the principle of lead time bias which simply means extending the period of observation so that a woman with a screen detected cancer may enjoy a longer period of post diagnosis survival, whilst dying at the same time as they would have done had the tumour been allowed to appear clinically.

Finally let us look at the Canadian National Breast screening Study (CNBSS). This study, because of its negative findings, has come under the most intensive scrutiny, which I like to describe as 'ideological bias', because the data are tortured until forced to confess! It is quite extraordinary to my mind that whilst we accept the HIP study where close on 40% of the cancers were clinically detected we wish to exclude the CNBSS where

the prevalent and incident rounds had mammographic detection rates and case survival rates as good as, if not better, than any other trial. Furthermore again we see a double standard. For treatment trials the meta-analyses include all properly randomized studies, many of which, by the play of chance, will include results in the 'wrong' direction but that doesn't mean sending in the thought police to beat up these miscreants!¹⁰

For example I find the following statement extremely distasteful:

the CNBSS log books were altered, but document experts found no evidence of a deliberate attempt to conceal the alterations. That seems weak: among other things, randomization could have been subverted simply by changing the order in which names were entered into the logbooks.

Why on earth would they wish to subvert the study, who's undeclared conflict of interest would have been served by this scientific misconduct? I hope that Freedman *et al.* have consulted their lawyers, because if I were facing that accusation my libel lawyer would already be sharpening up his pencil.

Like many screening advocates they finish their paper by saying 'it is time to move on'. This is a short hand way of trying to stifle the voice of irritating people like myself who continue to challenge their ideological beliefs. They qualify this statement by stating that '*some questions may remain*'. You betcha! Like, even if we accept their efficacy argument they haven't begun to address the clinical utility argument based on a harm-benefit analysis where the judge is the woman herself and not the agents of the state.¹¹

References

- 1 The ATAC Trialists' Group. Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early breast cancer: first results of the ATAC randomised trial. *Lancet* 2002;**359**:2131-39.
- 2 www.consort-statement.org. Date accessed: 5 August 2003.
- 3 Curevo LG, Clarke M. Balancing benefits and harms in health care. We need to get better evidence about harms. *BMJ* 2003;**327**:65-66.
- 4 Humphrey LL, Helfand M, Benjamin KS, Chan MS, Woolf SH. Breast cancer screening: a summary of evidence for the US Preventive Services Task Force. *Ann Intern Med* 2002;**137**:347-60.
- 5 Writing group for Women's Health Initiative investigators. Risks and benefits of estrogen and progestin in healthy postmenopausal women. Principle results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;**288**:321-33.
- 6 Freedman DA, Petitti DB, Robins JM. On the efficacy of screening for breast cancer. *Int J Epidemiol* 2004;**33**:43-55.
- 7 Tabar L, Yen M-F, Vital B, Chen H-H T, Smith R, Duffy SW. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *Lancet* 2003;**361**:1405-10.
- 8 NHS Breast Screening Programme & British Association of Surgical Oncology Breast Group. An audit of screen detected breast cancers for the year of screening April 1999 to March 2000. NHS cancer screening programmes 16 May 2001.
- 9 Douek M, Baum M. Mass breast screening: is there a hidden cost? *Br J Surg* 2003;**90**:44-45.
- 10 Early breast cancer trialists collaborative Group. (EBCTG) Effects of tamoxifen and cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28 896 women. *N Engl J Med* 1988;**319**:1681-92.
- 11 Thornton H, Edwards A, Baum M. Women need better information about routine mammography. *BMJ* 2003;**327**:101-03.