

Mammographic screening: evidence from randomised controlled trials

H. J. de Koning*

Department of Public Health, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, The Netherlands

Received 23 September 2002; accepted 3 April 2003

Background: All randomised breast cancer screening trials have shown a reduction in breast cancer mortality in the 'invited for mammography' screening arm compared with the 'control arm' for women aged 50 years and older at randomisation (overall 25%). However, individually published point estimates differ and concern has been raised about methodological quality and outcome measures.

Materials and Methods: Review of the evidence on breast cancer mortality reduction and discussion of the causes of difference in point estimates in the five Swedish and Canadian trials. A summary of the prerequisites for methodological quality and its available evidence from the trials is given. Data to support breast cancer mortality as a correct outcome measure are presented.

Results: There is no reason not to use breast cancer mortality as an outcome measure for trials intended to reduce breast cancer mortality, both from a clinical and a methodological point of view. Everything possible was performed in these trials in order to determine this outcome measure as accurately as possible. The fact that a few of the trials showed a relatively large breast cancer mortality reduction and others far lower reduction rates is irrelevant, if one does not consider the background situation in the region before the trial started, the design of the trial or quality of screening.

Conclusions: There seems no reason to change or halt the current nation-wide population-based screening programmes. Nor is there any justifiable reason for negative reports towards women or professionals.

Key words: breast cancer, randomised trials, screening

Introduction

Randomised screening trials are bothersome. It takes ages to come to an answer, and these need to be large-scale projects to be able to answer the questions [1]. A sufficient number of persons in the study arm have to be screened [2], but if randomisation has been carried out before consent (which is possible, depending on the local country-specific regulations), attendance for screening may be (too) low in practice [3]. If randomisation is carried out only after informed consent, this might affect the attitude of persons randomised to the control arm and lead to contamination, or it may lead to selection of enrolled persons in the first place. 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 [4].

Still, such long-term large-scale randomised screening trials are crucial, and there is no second-best option. Individual randomisation leads to comparable groups that can be traced for follow-up and evaluation and the final outcome can be linked directly to screening and intermediate outcome measures. If there is variation

in screening intensity, this sometimes can be evaluated. There are enough examples where the lack of randomised trials hamper present-day evaluations of service screening practice [5–7].

In breast cancer, both clinical and laboratory work support the theory of breast cancer screening. In the past, women treated by surgery for smaller tumours had a lower chance of distant metastases after long follow-up [8], and empirical work showed that treatment before a critical number of blood vessels had formed might prevent occurrence of metastases [9].

At the request of the Danish Institute for Health Technology Assessment, Danish researchers analysed and summarised all available literature on the nine randomised breast cancer screening trials performed in the past, which was then (partly) approved by the Cochrane group for publication in their library [10]. The breast cancer screening trials in question were all initiated during the period 1963–1982, and were concluded at least 10 years ago. In a more narrow sense, the review did not include any of the many breast screening programmes currently ongoing in Europe. All trials differed in age categories and number of women enrolled, starting year and place, and the way women were enrolled or selected. If we confine ourselves to the group of women aged 50 years and older at randomisation, as this is the group that is invited for screening in every centrally organised European screening programme, rather than women under the age of 50 years, all trials had published a reduction in breast cancer mortality in

*Correspondence to: Professor H. J. de Koning, Department of Public Health, Erasmus MC, University Medical Centre Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Tel: +31-10-408-7714; Fax: +31-10-408-9449; E-mail: h.dekoning@erasmusmc.nl

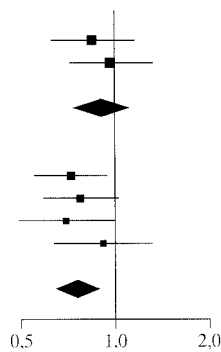


Figure 1. Relative risk (and 95% confidence interval) of dying from breast cancer in the mammography intervention arm compared with the 'control' arm for women aged $\geq 50/55$ years at the start of the study in the Cochrane review of six randomised breast cancer screening trials; 7-year follow-up [19]. The Canadian trial included an annual clinical breast examination by a nurse or physician in the control arm. Trial names in table 1.

the 'invited for mammography' screening arm compared with the 'control' arm, ranging from 3% to 36% [11, 12].

Causes of difference in point estimates on breast cancer mortality in the Swedish and Canadian trials

Two trials, one in New York [13] and one in Edinburgh [14], involved groups in two arms of the trials that were found to be unfit for comparison, either at the start or on analysis of the results. In Edinburgh, it ultimately emerged that breast cancer risk was not equally distributed between the intervention arm and the control arm of the study, a fact that was later extensively reported on by the trial's principal researchers [15]. Therefore, their results could not easily be utilised in estimating the effect of screening for breast cancer, nor were they consequently taken into consideration in either the recommendations for The Netherlands or those for the UK [16, 17]. The imbalance seen during the HIP trial in New York, on the other hand, came as a surprise. However, an essential fact is that the outcomes of the HIP trial were disregarded in the decision-making around many national screening programmes, as it was universally felt that the point estimate, dating from the 1960s, could have little relevance in the 1990s in Europe [18].

The entire discussion as to whether or not a reduction in breast cancer mortality could be established as a result of screening in women aged ≥ 50 years is therefore limited to the five Swedish studies and the Canadian study, comprising a total of 356 000 women. Figure 1 and Table 1 show the relative risk (RR) of women aged ≥ 50 years at the start of the trials (Malmö; age ≥ 55 years) dying of breast cancer in the intervention arm ('mammography screening') compared with the 'control' arm of these trials, summarised in the official Cochrane review [19]. An overall 25% (11–38%) statistically significant reduction in breast cancer mortality was seen in the intervention arms compared with the control arms. The overview shows a 12% reduction in breast cancer mortality in the Malmö and Canada trials (RR = 0.88) and a 31%

Table 1. Relative risk (RR) of women aged ≥ 50 years at the start of the trials dying of breast cancer in the intervention arm compared with the control arm

| Trial quality | RR with 95% CI |
|---------------|------------------|
| Medium | |
| Malmö | 0.80 (0.51–1.24) |
| Canada | 0.97 (0.62–1.52) |
| Subtotal | 0.88 (0.64–1.20) |
| Low | |
| Kopparberg | 0.63 (0.43–0.93) |
| Östergötland | 0.70 (0.47–1.04) |
| Stockholm | 0.59 (0.36–0.98) |
| Göteborg | 0.90 (0.53–1.54) |
| Subtotal | 0.69 (0.55–0.86) |

CI, confidence interval.

reduction in breast cancer mortality (RR = 0.69) in the four other Swedish trials. All the trials show a breast cancer mortality reduction for this age group, but only a few were on a large enough scale individually to reveal a significant reduction.

The fact that a few of the trials showed a relatively large breast cancer mortality reduction and others far lower reduction rates is irrelevant, if one does not consider the background situation in the region before the trial started, the design of the trial or quality of the screening. The difference between the arms (or before and after service screening started) is a result of these and other factors such as attendance figures, detection rates, technical quality and referral rates. We have already demonstrated that the variation between the Malmö trial and the Kopparberg/Östergötland trial could have arisen solely as a consequence of differences in screening policy (i.e. screening interval, attendance rate, follow-up years). In fact, based only on trial design characteristics, we would have expected breast cancer mortality reduction to range from 24% to 32% in the five Swedish trials [20]. The observed reductions, as published in 1993, for example, represent follow-up years differing from only 6.2 years for Göteborg to 11.8 years for Malmö [21], and are likely to influence results [20, 21].

The Canadian trial was the only trial that was uniquely different in design. The women assigned to the control groups in the Swedish trials were offered no examination, while in the Canadian trial, the women allotted to the control group annually received an extensive clinical breast examination, performed by a specially instructed and trained nurse or physician [22]. This yielded detection rates in the control arm rivalling those found in some decentral mammography screening programmes in Europe [23]. In short, one part of the explanation for the relatively small difference between the two arms in the Canadian trial (a mere 3%) is explained by the effective screening carried out in the control arm. This reported point estimate could therefore not be pooled in the same way with the Malmö trial to compare the effect of mammography screening to a no-screen situation, as the effect has been diluted.

Prerequisites for methodological quality of the trials

Of course, prerequisites are: (i) a correct randomisation, preferably at the individual level or with small homogenous blocks; (ii) post-randomisation exclusions for already existing cancer cases before randomisation (that are legitimate) being based on the same type of information in both arms; (iii) adequate follow-up of deaths in both arms being possible; and (iv) a blinded review of cause of death. The randomisation methods in the aforementioned six trials seem to have been justified: they have been based on computer ranking, printed allocation lists and day of birth as unit of allocation or an official clerk using the 'coin-method' to randomly allocate geographical areas at the level of village or town, in the two largest Swedish trials, Kopparberg and Östergötland, thereby taking socio-economic variations into account [24]. Although it was shown that the women living in the villages randomised to the intervention arm tended on average to be older than the women living in the villages allotted to the control arm, it was recently demonstrated that the latter strategy has been successful and that the heterogeneity between the clusters and within the strata of the clusters was very small and moreover that this barely affected the final outcomes [25].

All the Swedish trials were population-based, which means that in principle, a population list of women was compiled in advance and randomised. Obviously, it is possible and even likely that this included women in whom breast cancer had been diagnosed prior to the randomisation date. These women will have to be excluded later, for example on the basis of information derived from the cancer registry. This was the case for a few hundred women in each Swedish trial. Equal exclusions were reported for Östergötland, Malmö and Göteborg, and no age differences were seen between arms in Stockholm. The Canadian trial was the sole trial conducted on a volunteer basis, i.e. on women who approached the centres on their own initiative. Hence, this trial had almost no exclusions, as women already diagnosed with breast cancer are unlikely to join, or will not be randomised if they do volunteer. The Canadian trial is therefore more likely to fulfil this exclusion criterion, while at the same time saying nothing about a methodological error or quality criterion. The discussion concerning age differences between arms constituting the criterion for correct randomisation [26] has proven to be obsolete and wrong.

Breast cancer mortality as outcome measure

A group of independent clinicians carried out a systematic and blind assessment of the deaths of all women in the five Swedish trials in whom breast cancer had been diagnosed (regardless of how this had been diagnosed and the arm to which they had been allotted), to determine the cause of death. Blind, in this context, refers to the fact that the reviewers did not know to which arm the women belonged, nor in what way the diagnosis had been made. Table 2 shows the number of deaths attributable to breast cancer as determined by the Swedish Bureau for Statistics in the arms of all five Swedish trials, and the figures derived by the independent

Table 2. Number of breast cancer deaths, person years and relative risk (RR) (and 95% CI) in invited group (IG) and control group (CG) for 'breast cancer as underlying cause of death' and 'breast cancer present at death', independent and blind assessment by End Point Committee (EPC), and for 'breast cancer as underlying cause of death' according to Official Statistics Sweden, by age of the women at randomisation in five Swedish randomisation trials [30]

| Age at randomisation, years | Person years (×1000) | | Independent and blind assessment (EPC) | | | | | | Official Statistics Sweden | | | | | |
|-----------------------------|----------------------|------|--|-----|--------|--------------------------------|-----|--------|--|-----------|--------|--|------|-----------|
| | | | Breast cancer as underlying cause of death | | | Breast cancer present at death | | | Breast cancer as underlying cause of death | | | Breast cancer as underlying cause of death | | |
| | IG | CG | Numbers | RR | 95% CI | Numbers | RR | 95% CI | Numbers | RR | 95% CI | Numbers | RR | 95% CI |
| 50–59 | 540 | 454 | 151 | 174 | 0.72 | 0.58–0.90 | 153 | 181 | 0.71 | 0.58–0.88 | 153 | 167 | 0.76 | 0.61–0.95 |
| 60–69 | 372 | 271 | 130 | 138 | 0.69 | 0.54–0.88 | 147 | 144 | 0.75 | 0.59–0.95 | 125 | 133 | 0.69 | 0.54–0.88 |
| 40–74 | 1430 | 1139 | 418 | 425 | 0.77 | 0.67–0.88 | 440 | 442 | 0.79 | 0.69–0.90 | 419 | 409 | 0.80 | 0.70–0.92 |

committee. While the absolute figures obviously differ to some extent, the differences in the reduction of breast cancer mortality were not significantly affected.

Our own studies have shown that once women develop breast cancer metastases, sadly, around 95% will go on to die of this disease [27, 28]. The clinical course of metastatic breast cancer nearly always ends in a fatal outcome; death from other causes is rare. The suggestion that breast cancer mortality is an unreliable outcome measure therefore fails to hold water, both from a clinical and a methodological point of view.

Evaluation service screening

Various randomised breast cancer screening trials in the past have demonstrated that a reduction in breast cancer mortality as a result of screening is certainly feasible for women aged ≥ 50 years. It is now time to move on to the evaluations of service screening programmes. In making the decision to launch a screening programme, we estimated that by 1999, around half of the maximum effect would be visible. It is therefore premature to give any kind of substantiated opinion based on national mortality figures as to whether the present breast cancer screening programme is sufficiently or insufficiently effective. The actual impact of the national screening programme in The Netherlands cannot yet be ascertained. Preparations are currently in progress by the National Evaluation Team for Breast Cancer Screening to gain more insight into this aspect by means of a detailed mortality review. The population screening programme started gradually; it was not until 1993 that over half of the target group was invited for initial screening, yet by 1997 all the women in the age group between 50 and 69 years had been invited for screening at least once.

Conclusions

Cancer screening has always been a controversial topic, partly because it concerns the examination of people who are, in principle, healthy, for whom the health gain must be absolutely undisputed. Evidence of this health gain, however, relates to the group as a whole; at the individual level it is impossible to determine who will benefit and who will suffer harm. There is a highly delicate balance between favourable and unfavourable effects [29]. This is why extensive analyses of the expected health gain, other favourable and unfavourable side-effects and the investment required are carried out prior to introducing large-scale screening programmes of this kind.

There is no reason not to use breast cancer mortality as an outcome measure for trials intended to reduce breast cancer mortality. Everything possible was performed in these trials in order to determine this outcome measure as correctly as possible.

It is more than likely that the reduction in breast cancer mortality in the various countries may indeed be attributed to the screening programmes, although it is still too early for a well-founded, scientific opinion in this respect. It goes without saying that evaluation of breast cancer mortality during the next 5 years will be crucial.

There seems no reason to change or to halt the current screening programmes, for example in The Netherlands. Nor is there any justifiable reason for negative reports towards women or professionals.

References

1. de Koning HJ, Auvinen A, Berenguer Sanchez A et al. Large-scale randomized prostate cancer screening trials: program performances in the European randomized screening for prostate cancer, and the prostate, lung, colorectal and ovary cancer trial. *Int J Cancer* 2002; 97: 237–244.
2. de Koning HJ, Liem M, Baan CA et al. Prostate cancer mortality reduction by screening; power and time frame with complete enrollment in the European randomised screening for prostate cancer (ERSPC) trial. *Int J Cancer* 2002; 98: 268–273.
3. Labrie F, Candas B, Dupont A et al. Screening decreases prostate cancer death: first analysis of the 1988 Quebec prospective randomized controlled trial. *Prostate* 1999; 38: 83–91.
4. Beemsterboer PMM, Kranse R, de Koning HJ et al. Changing role of 3 screening modalities in the European randomized study of screening for prostate cancer (Rotterdam). *Int J Cancer* 1999; 84: 437–441.
5. Sasieni PD, Cuzick J, Lynch-Farmery E. Estimating the efficacy of screening by auditing smear histories of women with and without cervical cancer. The National Co-ordinating Network for Cervical Screening Working Group. *Br J Cancer* 1996; 73: 1001–1005.
6. Läärä E, Day NE, Hakama M. Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. *Lancet* 1987; 30: 1247–1249.
7. Henschke CI, McCauley DI, Yankelevitz DF et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999; 10: 99–105.
8. Koscielny S, Tubiana M, Le MG et al. Breast cancer: relationship between the size of the primary tumour and the probability of metastatic dissemination. *Br J Cancer* 1984; 49: 709–715.
9. Horak ER, Leek R, Klenk N et al. Angiogenesis, assessed by platelet/endothelial cell adhesion molecule antibodies, as indicator of node metastases and survival in breast cancer. *Lancet* 1992; 340: 1120–1124.
10. Cochrane group. library <http://www.update-software.com/cochrane/reviews>.
11. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: I and II. *Can Med Assoc J* 1992; 147: 1459–1488.
12. Frisell J, Eklund G, Hellstrom L et al. Randomized study of mammography screening—preliminary report on mortality in the Stockholm trial. *Breast Cancer Res Treat* 1991; 18: 49–56.
13. Shapiro S, Strax P, Venet L. Evaluation of periodic breast cancer screening with mammography. Methodology and early observations. *JAMA* 1966; 195: 731–738.
14. Roberts MM, Alexander FE, Anderson TJ et al. The Edinburgh randomised trial of screening for breast cancer: description of method. *Br J Cancer* 1984; 50: 1–6.
15. Alexander F, Roberts MM, Lutz W, Hepburn W. Randomisation by cluster and the problem of social class bias. *J Epidemiol Community Health* 1989; 43: 29–36.
16. Forrest P. Breast cancer screening. Report to the Health Ministers of England, Wales, Scotland & Northern Ireland. London: Her Majesty's Stationery Office 1986.
17. van der Maas PJ, van Ineveld BM, van Oortmarssen GJ et al. Costs and effects of breast cancer screening. Interim report + annex. Department of Public Health. Rotterdam: Erasmus MC 1987.
18. de Koning HJ, Boer R, van der Maas PJ et al. Efficacy of mass screening for breast cancer; reduced mortality nationally and internationally. *Ned Tijdschr Geneesk* 1990; 134: 2240–2245.

19. Olsen O, Gøtzsche PC. Screening for breast cancer with mammography (Cochrane Review). In *The Cochrane Library*, Issue 4. Oxford: Update Software 2001.
20. de Koning HJ, Boer R, Warmerdson PG et al. Quantitative interpretation of age-specific mortality reductions from the Swedish Breast Cancer Screening trials. *JNCI* 1995; 87: 1217–1223.
21. Miettinen OS, Henschke CI, Pasmantier MW et al. Mammographic screening: no reliable supporting evidence? *Lancet* 2002; 359: 404–405.
22. Baines CJ, Miller AB, Bassett AA. Physical examination. Its role as a single screening modality in the Canadian National Breast Screening Study. *Cancer* 1989; 63: 1816–1822.
23. Beemsterboer PMM, Koning HJ de, Warmerdam G et al. Prediction of the effects and costs of breast-cancer screening in Germany. *Int J Cancer* 1994; 58: 623–628.
24. Tabár L, Gad A, Akerlund E et al. Screening for breast cancer in Sweden. A randomised controlled trial. In Logan WW, Muntz EP (eds): *Reduced Dose Mammography*. New York: Masson 1979; 407–414.
25. Nixon R, Prevost TC, Duffy SW et al. Some random-effects models for the analysis of matched-cluster randomised trials: application to the Swedish two-county trial of breast-cancer screening. *J Epidemiol Biostat* 2000; 5: 349–358.
26. Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000; 355: 129–134.
27. de Koning HJ, van Ineveld BM, de Haes JCJM et al. Advanced breast cancer and its prevention by screening. *Br J Cancer* 1992; 65: 950–955.
28. Richards MA, Braysher S, Gregory WM et al. Advanced breast cancer: use of resources and cost implications. *Br J Cancer* 1993; 67: 856–860.
29. de Koning HJ. Assessment of nationwide cancer-screening programmes. *Lancet* 2000; 355: 80–81.
30. Nyström L, Larsson LG, Rutqvist LE et al. Determination of cause of death among breast cancer cases in the Swedish randomized mammography screening trials. A comparison between official statistics and validation by an endpoint committee. *Acta Oncol* 1995; 34: 145–152.