

Cochrane review on screening for breast cancer with mammography

Ole Olsen, Peter C Gøtzsche

In 2000, we reported that there is no reliable evidence that screening for breast cancer reduces mortality. As we discuss here, a Cochrane review has now confirmed and strengthened our previous findings. The review also shows that breast-cancer mortality is a misleading outcome measure. Finally, we use data supplemental to those in the Cochrane review to show that screening leads to more aggressive treatment.

Lancet 2001; **358**: 1340-42

See [Commentary](#)

We previously assessed the results of the seven randomised trials of screening mammography, and concluded that screening is unjustified because there is no reliable evidence that it reduces mortality.¹ We reassessed this finding in a Cochrane review² in which we paid close attention to the standard dimensions of methodological quality of trials: the randomisation method, baseline comparability, exclusions after randomisation, and unbiased assessment of outcome (see protocol for the Cochrane review [issue 3, 2001, Cochrane Library]). Additionally, we noted whether early introduction of screening in the control group had occurred. Details of the trial assessments are presented in our review.² On the basis of these assessments, we classified the quality of the available trial data into four groups: high, medium, poor, and flawed.

We found that the results confirmed and strengthened our original conclusion. No trial data were of high quality, two were of medium quality (Malmö and Canada), three were of poor quality (Two-County, Stockholm, and Göteborg), and two were flawed (New York and Edinburgh). The review provided evidence that assessment of cause of death is unreliable and biased in favour of screening. Even when endpoint committees masked to group assignment were used, uncertain causes of death were significantly more commonly ascribed to breast cancer than to other causes in the control group. The credibility of this finding is supported by another meta-analysis, which showed that radiotherapy reduces local recurrence by two-thirds.³ Treatment of early cancers by tumourectomy and radiotherapy might increase the likelihood that deaths among screen-detected breast cancer cases will be misclassified as deaths from other causes,³ particularly other cancers.² We noted that the two trials with medium-quality data failed to find an effect of screening on deaths ascribed to any cancer, including breast cancer (relative risk 1· 02 [95% CI 0· 95-1· 10]). The estimate for the trials with poor-quality data was similar (1· 00 [0· 91-1· 10]). Furthermore, the greater use of radiotherapy in screened women than in controls¹ is expected to increase overall mortality because of cardiovascular adverse effects.³ These deaths were not counted as deaths related to screening in the trials we assessed.

The main outcome measure in the screening trials was breast-cancer mortality. This choice seems rational, since larger trials would be needed to show an effect on overall mortality. However, we showed that the assumption that a demonstrated effect on breast-cancer mortality can be translated into a reduction in overall mortality rests on suppositions that are not correct.² The only reliable mortality estimates are therefore those for overall mortality. The relative risk of overall mortality was 1· 00 (0· 96-1· 05) in the two trials of highest methodological quality (figure).² The Swedish trialists have recently reported an updated mortality estimate for the four Swedish trials:⁴ this estimate was also 1· 00 (0· 98-1· 02) after adjustment for imbalances in age that had occurred despite attempts at randomisation.^{1,2} Thus, although the trials were underpowered for all-cause mortality, the reliable evidence does not indicate any survival benefit of mass screening for breast cancer.



All-cause mortality in medium-quality screening trials after 13 years

*Fixed-effects model.

In our previous paper,¹ we divided the trials into two groups on the basis of methodological quality. We reported that the effect estimate for breast cancer mortality in the two best trials was significantly different from that for the five poor-quality trials, which is a sign that something is wrong. In our latest review, we therefore omitted the trials from New York and Edinburgh from the analysis of the poor-quality trials, since

they are flawed.² However, there was still a significant difference between the two estimates for breast-cancer mortality. The two best trials failed to find an effect of screening on deaths ascribed to breast cancer (relative risk 0.97 [0.82-1.14] after 13 years, whereas the three remaining trials with poor-quality data found a marked effect (0.68 [0.58-0.78]; $p=0.001$ for the difference between the two effect estimates). Given the strong heterogeneity, results from the different quality groups should not be combined.

The largest effects on breast-cancer mortality were reported in trials that had long intervals between screenings (Two-County trial), that invited many women to only two or three screenings (Two-County and Stockholm trials), that started systematic screening of the control group after 3-5 years (Two-County trial, Göteborg trial, and Stockholm trial) and that had poor equipment for mammography (New York trial). This surprising situation suggests that differences in reported effects between the trials are related to the methodological quality of the trials and not to the quality of the mammograms or the screening programmes.²

We have also confirmed, with additional data (see www.thelancet.com), which the editors of the Cochrane Breast Cancer Group have elected to defer from publication until further editorial review has been completed, our earlier finding¹ that screening leads to more aggressive treatment, increasing the number of mastectomies by about 20% and the number of mastectomies and tumourectomies by about 30%. The greater use of surgery was not merely an initial phenomenon caused by the tumours detected at the prevalence screen, but seemed to persist. The increased mastectomy rate in the trials might be higher than in current practice, since there has been a general policy change towards fewer mastectomies. However, screening identifies some slow-growing tumours that would never have developed into cancer in the women's remaining lifetimes, as well as cell changes that are histologically cancer but biologically benign. Furthermore, carcinoma in situ does not always develop into invasive cancer, but since these early lesions are often diffuse, women are sometimes treated by bilateral mastectomy. Therefore, the increase in surgery rates could also be an underestimate, since reoperations and operations in the contralateral breast seemed not to have been included. Furthermore, "better" diagnostic methods--eg, better mammograms--could lead to additional overtreatment because of detection of even more early or questionable lesions. Quality assurance programmes could possibly reduce the surgical activity to some degree, but the problem cannot be avoided.

Our earlier report¹ has been criticised,^{5,6} especially for its emphasis on imbalances in baseline variables. However, the main reason for the ongoing controversy is probably that our opponents keep referring to the criticisms of our paper without referring to our reply.⁷ Furthermore, they seem to have ignored this sentence in our paper: "Our analyses focused on age as a marker for imbalance as this was the only baseline information we had available for the Swedish trials".¹ We have not postulated that the baseline imbalances per se caused the inflated effect, but we used the imbalances as markers of poor trial methodology⁷--an approach that led us to new important information about the trials.² Contrary to what the critics assert,⁶ the fact that there was no age imbalance in the two best trials was confirmed in the correspondence that followed our *Lancet* paper, and we believe that all relevant criticism has now been addressed in our review.²

We have provided detailed evidence on the mammography screening trials, and hope that women, clinicians, and policy-makers will consider these findings carefully when they decide whether or not to attend or support screening programmes. Any hope or claim that screening mammography with more modern technologies than applied in these trials will reduce mortality without causing too much harm will have to be tested in large, well-conducted randomised trials with all-cause mortality as the primary outcome.

This study was funded by the Danish Institute for Health Technology Assessment. The review is available at <http://image.thelancet.com/lancet/extra/fullreport.pdf>

1 Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000; **355**: 129-34. [[Text](#)]

2 Olsen O, Gøtzsche PC. Screening for breast cancer with mammography. In: Cochrane Library, issue 4. Oxford: Update Software (in press).

3 Early Breast Cancer Trialists' Collaborative Group. Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 2000; **355**: 1757-70. [[Text](#)]

4 Nyström L. Assessment of population screening: the case of mammography. Umeå UmeåUniversity Medical Dissertations, 2000 (thesis).

5 Wald N. Populist instead of professional. *J Med Screen* 2000; **7**: 1.

6 Duffy SW. Interpretation of the breast screening trials: a [commentary](#) on the recent paper by Gøtzsche and Olsen. *Breast* 2001; **10**: 209-12. [[PubMed](#)]

7 [Gøtzsche PC, Olsen O. Screening mammography re -evaluated.](#) *Lancet* 2000; **355**: 752.