Is screening for breast cancer with mammography justifiable?

Peter C Gøtzsche, Ole Olsen

Summary

Background A 1999 study found no decrease in breast-cancer mortality in Sweden, where screening has been recommended since 1985. We therefore reviewed the methodological quality of the mammography trials and an influential Swedish meta-analysis, and did a meta-analysis ourselves.

Methods We searched the Cochrane Library for trials and asked the investigators for further details. Meta-analyses were done with Review Manager (version 4.0).

Findings Baseline imbalances were shown for six of the eight identified trials, and inconsistencies in the number of women randomised were found in four. The two adequately randomised trials found no effect of screening on breast-cancer mortality (pooled relative risk 1·04 [95% CI 0·84–1·27]) or on total mortality (0·99 [0·94–1·05]). The pooled relative risk for breast-cancer mortality for the other trials was 0·75 (0·67–0·83), which was significantly different (p=0·005) from that for the unbiased trials. The Swedish meta-analysis showed a decrease in breast-cancer mortality but also an increase in total mortality (1·06 [1·04–1·08]); this increase disappeared after adjustment for an imbalance in age.

Interpretation Screening for breast cancer with mammography is unjustified. If the Swedish trials are judged to be unbiased, the data show that for every 1000 women screened biennially throughout 12 years, one breast-cancer death is avoided whereas the total number of deaths is increased by six. If the Swedish trials (apart from the Malmö trial) are judged to be biased, there is no reliable evidence that screening decreases breast-cancer mortality.

Lancet 2000; **355:** 129–34
See Commentary page xxxx

Introduction

After heated controversy, there now seems to be general acceptance that the benefit of screening for breast cancer with mammography has been well documented.[1] Large randomised trials, including a total of half a million women, have been carried out in New York, USA;[2] Edinburgh, Scotland;[3] Canada;[4,5] and Malmö,[6] Kopparberg,[7] Östergötland,[7] Stockholm,[8] and Göteborg[9] in Sweden. A meta-analysis of an update of the five Swedish trials, which used data from individual patients, was particularly influential. It showed that screening lowered mortality from breast cancer by 29% in women aged 50–69 years.[10]

The findings of a 1999 epidemiological study were therefore surprising. It found no decrease in breast-cancer mortality in Sweden,[11] where screening has been recommended since 1985. The observed decrease in number of deaths from breast cancer was 0·8% (not significant), whereas the expected decrease was 11%. Although that study can be criticised,[12,13] it raises once again the issue of the reliability of the evidence that screening is effective.

We therefore reviewed the methodological quality of the mammography trials and the Swedish meta-analysis, and did a meta-analysis ourselves. We focused on the three most important sources of bias in randomised trials: suboptimum randomisation methods, lack of masking in outcome assessment, and exclusion after randomisation. We paid special attention to the quality of the randomisation, since bias caused by suboptimum randomisation methods can be larger[14,15] than the treatment effects that might be detected if a screening programme is beneficial.

Methods

We searched the Cochrane Library with the terms "breast-neoplasms/all" or "breast next cancer" and "screening" and "mammography" and extended the search with authors' names and other terms as appropriate to capture updates of the trials. When necessary, we asked the investigators for details about the randomisation method, in particular whether the assignment process was concealed so that no-one could foresee which assignment the next cluster or woman would get before actual recruitment. We also asked for baseline characteristics that could show whether the screening group was similar to the control group in terms of important prognostic factors such as age, symptoms at entry, family history of breast cancer, socioeconomic status, and previous examinations for breast cancer. We noted whether all randomised women had been accounted for in the results and whether the cause of death had been assessed by a panel unaware of screening status. We also sought data on the morbidity associated with screening, defined as reported events that had occurred in at least 100 women.

Meta-analyses were done with Review Manager (version 4.0; available from http://www.cochrane.dk; accessed on Dec 20, 1999). A fixed-effects model was used unless the test for heterogeneity gave p<0·10; 95% CIs are presented.

Results

Randomisation methods and exclusions

In the New York trial, pairs of women were matched and the pairs were randomised.[16] The allocation method is not clear—"every nth woman was placed in the study group, the paired (n+1) woman in the control group".[16] Because of the matching in pairs, the number of randomised women should be exactly the same in the study group and in the control group. This was not the case, and the number of women is unclear. It has been described as "about 31 000",[16] 30 000,[17] 30 131,[2] 31 092,[18] and 30 239[19,20] allocated to the study group, and 30 756,[20] 30 765,[19] and 30 565[2,16] allocated to the control group. There was also an important imbalance in exclusions after randomisation. Women were excluded if breast cancer had been diagnosed before entry to the trial, and this

Nordic Cochrane Centre, Rigshospitalet, Department 7112, Blegdamsvej 9, 2100 Copenhagen Ø, Denmark (P C Gøtzsche MD, O Olsen MSc)

Correspondence to: Dr Peter C Gøtzsche (e-mail: p.c.gotzsche@cochrane.dk)

status was more completely ascertained for the screened women; thus, the final study cohort was smaller than the control cohort (30 131 *vs* 30 565).[2,16] This difference introduced bias in favour of the screening group. Close similarity between the study and control groups has been claimed,[16,21] but in the table of seven selected characterisics presented in justification for this claim, we calculated imbalances for previous lump in the breast ($p<0·0001$), menopause ($p<0·0001$), and education ($p=0·05$); there were no differences for age, religion, marital status, or pregnancies. These findings are incompatible with an adequate randomisation.

The allocation method of the Edinburgh trial is poorly described; 87 general practices were cluster randomised,[22] but the allocation was later changed for three of them.[23] The screening and control groups differed substantially at baseline; only 26% of the women in the control group were in the highest socioeconomic stratum, compared with 53% in the screening group.[22] Thus, the randomisation method was grossly inadequate, even for a cluster analysis.

In the Canadian trial, women were randomised individually.[24] Names were entered successively on allocation lists, in which the intervention was noted on each line. The randomisation could therefore be subverted. However, checking of whether this had happened was also possible, and a thorough review concluded that there could not have been enough cases of such subversion to affect the reported results.[25] Moreover, the two compared groups were similar at baseline in terms of self-reported symptoms, including lump, family history of breast cancer, marital status, livebirths, menopause, education, and place of birth.[26,27] We found no data on the age distribution.

In the Malmö trial,[6] women in each birth-year cohort were randomly arranged according to a computer program, and those on the first half of the lists were invited for screening (Ingvar Andersson, personal communication). Thus, the allocation method was apparently adequately concealed. No baseline data are available, but we estimated from the other Swedish trials that the mean age was similar in the two groups.

A sort of continuation of that trial, called Malmö Mammographic Screening Trial II,[28] has been published in brief; it was randomised and had death from breast cancer as the endpoint, but it did not have a formal protocol, and because of an administrative error, all women born in 1934 were included in the screening group (Ingvar Andersson, personal communication). Because the report mixes follow-up data from a subgroup of the original trial with data from this new cohort, and since some women were not randomised, the published data cannot be included in a meta-analysis. No baseline data are available.

In the Stockholm trial,[8] randomisation was according to date of birth; women born on days 11–20 of any month constituted the control group. The number of randomised women is not clear. The number of controls is given as "c. 20 000" in an early report,[29] and as 19 943 in the final report.[8] There is a substantial discrepancy between the numbers in the final report and the meta-analysis of the Swedish trials[10] in which the number of randomised women fell from 40 318 to 38 525 (a decrease of 4·5%) in the screening group, but increased from 19 943 to 20 651 (a rise of 3·6%) in the control group. This inconsistency cannot be explained by the curious fact that women born on day 31 of any month were excluded after randomisation despite being offered mammography "to simplify the numerical comparisons",[30] since that approach led to a study group size of 39 164 women. We cannot understand how the number of randomised women in the control group can increase. Some 40-year-old women were excluded from the meta-analysis, which was based on age at randomisation and not on birth-year cohorts as most of the trials had used, but this exclusion would lead to a decrease as it did for the other three Swedish trials for which we could check the numbers (Malmö −1·9% *vs* −1·9%,[6] Kopparberg −1·3% *vs* −2·0%,[31] and Östergötland −0·2% *vs* −0·7%[7]). We calculated from a table divided into five age categories[30] that the study women in Stockholm were, on average, 0·18 years younger than the control women ($z=2·73$, $p=0·006$, Mann-Whitney test). This imbalance at baseline indicated that the randomisation method was inadequate.

In Göteborg, randomisation was partly by day-of-birth cluster (18% of participants) and partly individual.[9] We calculated from a table divided into 11 age categories[9] that the study women were, on average, significantly younger than the control women by 0·09 years ($z=2·39$, $p=0·02$), which shows that the randomisation method may have been inadequate.

Cluster randomisation was used in Kopparberg and Östergötland.[32] The population in these counties was divided into 19 blocks which were further divided into two or three groups on unspecified criteria. These groups were then randomised. We were unable to find a description of the randomisation method. In Nyström and colleagues' meta-analysis, the cluster randomisation method was said not to have introduced bias.[10] However, the justification for this statement was a reference to an unpublished lecture.[10] The meta-analysis is unlikely to have taken the clustering into account, since we obtained the same point estimate and the same narrow CI for breast-cancer mortality as in the meta-analysis when we based our analysis on individual women. We therefore used women as the statistical unit and calculated from a table divided into eight age categories[31] that the study women in Kopparberg were, on average, 0·45 years older than the control women ($z=5·50$, $p<0·0001$). There was also an imbalance in Östergötland ($z=4·04$, $p<0·0001$), the study women being 0·27 years older than the control women.[7] The number of randomised women (aged 40–74) is not clear: for example, the number in the study group in Östergötland has been reported as 39 034[32,33] and 38 491;[7,34] the total number of randomised women in the two trials has been reported as 134 867[32] and 133 065.[7,34]

Baseline data were not reported in the Swedish meta-analysis.[10] 3 years after the report was published in *The Lancet*, however, a report in a specialist journal stated that the mean age in the screened groups was 55·05 years compared with 54·54 years in the control groups.[35] Since the SD for age in the Swedish trials was 10 years,[7,31] the age difference was highly significant ($z=12·7$, $p=3\times10^{-37}$). This extremely skewed distribution is incompatible with the hypothesis that the women were distributed to the screening and control groups according to a truly chance procedure.

We estimated whether the Malmö trial had an imbalance at baseline like the other four Swedish trials. We used the number of women as reported in the meta-analysis and the mean ages as estimated above. We took account of the fact that women in Göteborg were

	Randomisation produced similar groups	Account of number of patients consistent
Malmö	Yes	Yes
Canada	Yes	Yes
Göteborg	No	Yes
Stockholm	No	No
Kopparberg	No	No
Östergötland	No	No
New York	No	No
Edinburgh	No	Yes

Table 1: **Mammography screening trials according to methodological quality**

randomly allocated to study and control groups in the approximate ratio of 1·2 in the 39–49-year age-group and 1·6 in the 50–59-year age-group.[9] We had no data on age for the 50–59-year group, but since the imbalance in age in the 39–49-year group was numerically small, we used a mean age of 54 for both study and control groups. For Malmö, we used 57 years as estimated mean age in the study group, similar to the Kopparberg and Östergötland trials.[7,31] This approach yielded a mean age in the study groups of 54·93 years, very close to the 55·05 years reported in the meta-analysis. Since the mean age in the control groups was 0·51 years lower, that in the Malmö control group was estimated to be 56·85 years. The difference of 0·15 years is not significant (z=1·53, p=0·13) which suggests that the randomisation method in Malmö was adequate. In summary, our findings suggest that only the trials from Malmö and Canada were unbiased (table 1).

Diagnosis of deaths from breast cancer

Knowledge of screening status may affect the judgment of cause of death. Masked assessment of cause of death was used only in the trials from Canada and Malmö, but in the Swedish meta-analysis[10] all deaths from breast cancer were assessed with masking of screening status. Deaths from breast cancer diagnosed before entry to the trial were generally excluded from analysis. Such exclusions can lead to bias when the first round of screening identifies cancer in women who have already noted a tumour in their breast if these women are subsequently excluded. The New York trial excluded more cancers in the screening group than in the control group.

All-cause mortality

The imbalance in age at baseline in the Swedish trials is important. Nyström and colleagues reported in a specialist journal[35] that the screened women had an increased risk of death (relative risk 1·05; 15 695 women died of 156 911 in the screening groups *vs* 11 887 of 125 866 in the control groups). Nyström and colleagues did not test whether this increased mortality was significant, nor did they give a CI. They argued that because breast-cancer mortality constitutes less than 5% of the total mortality, such an analysis "would require very large cohorts and is therefore impossible in practice".[35] We based our calculation on number of randomised women (the meta-analysis investigators had used person-years) and found a relative risk of 1·06 (95% CI 1·04–1·08, p<0·0001). The investigators adjusted their calculation for age, after which the relative risk was 1·00. In *The Lancet* report of the meta-analysis,[10] the investigators had included the same total numbers of deaths but reported only the age-adjusted risk without mentioning that an adjustment had been made or that there was an increased risk of death without adjustment.

	Number randomised		Number of deaths from breast cancer		Relative risk (95% CI)
	Screening	Control	Screening	Control	
Randomisation adequate					
Malmö[6]	21 088	21 195	63	66	0·96 (0·68–1·35)
Canada[5,27]	44 925	44 910	120	111	1·08 (0·84–1·40)
Total	66 013	66 105	183	177	1·04 (0·84–1·27)
Randomisation not adequate					
Göteborg[9]	11 724	14 217	18	40	0·55 (0·31–0·95)
Stockholm[8]	40 318	19 943	66	45	0·73 (0·50–1·06)
Kopparberg[7]	38 589	18 582	126	104	0·58 (0·45–0·76)
Östergötland[7]	38 491	37 403	135	173	0·76 (0·61–0·95)
New York[2]	30 131	30 565	153	196	0·79 (0·64–0·98)
Edinburgh[3]	22 926	21 342	156	167	0·87 (0·70–1·08)
Total	182 179	142 052	654	725	0·75 (0·67–0·83)

Table 2: **Relative risk of death from breast cancer in screened versus control groups**

The pooled relative-risk estimate for the two unbiased trials (Malmö and Canada) was 0·99 (0·94–1·05), which was very close to the estimate for Malmö alone (0·99 [0·93–1·05]), since that study reported 3586 deaths, compared with only 1147 in Canada (relative risk 1·08 [0·84–1·40]).

Mortality from breast cancer

The two trials with adequate randomisation methods and baseline comparability (table 1) had similar estimates for the relative risk of death from breast cancer with 95% CIs that overlapped substantially, showing lack of heterogeneity (table 2). The combined relative-risk estimate was 1·04 (0·84–1·27).

The six trials that had not been adequately randomised had more favourable outcomes with screening than these two trials, and their results were homogeneous (p=0·23 for test of heterogeneity). The pooled relative risk was 0·75 (0·67–0·83). This estimate is significantly different from that for the two adequately randomised trials (z=2·60, p=0·005).

If the Göteborg trial, which was the least biased trial of the six, was moved from the second group to the first, the relative-risk estimate changed little (0·94 [0·76–1·17]). However, since this change creates heterogeneity (p=0·08), this trial should probably not be moved. If all eight trials are analysed together (which would be inappropriate), heterogeneity is also introduced (p=0·05).

Morbidity

Total numbers of interventions were identified only in the trials from Malmö[6] and Stockholm.[29] Surgery was significantly more common in the screening groups for radical mastectomy (relative risk 1·23 [1·08–1·40]) and for mastectomy or lumpectomy (1·35 [1·20–1·52], as was radiotherapy (1·25 [1·04–1·50]). A similar tendency was seen in the Canadian trial, in which only surgery done within the framework of the trial was reported. In that trial, the proportion of benign findings in biopsy samples was two to four times higher in the mammography groups throughout the whole screening period.[5] We found no data from Edinburgh and New York and data only from the screened group for the other trials.

Discussion

The effect of screening programmes, if any, is small and the balance between beneficial and harmful effects is very delicate. It is therefore essential that such programmes are rigorously evaluated in properly randomised trials.

Unfortunately, the randomisation process failed to create similar groups in six of the eight trials of mammographic screening. Our analyses focused on age as a marker for imbalance, since this variable was the only baseline information we had available for the Swedish trials.

Cluster randomisation was used in several of the trials, but the number of clusters was insufficient, which is well illustrated by the Edinburgh trial.[22] The proportions of women in the highest socioeconomic stratum differed substantially between the screening and control groups, and, as expected, there was a pronounced relation between social group and total mortality, which may explain why total mortality was much lower in the screening group (relative risk 0·85 [0·79–0·92]). Attempts were made to remedy this shortcoming,[3] but adjustments cannot fully compensate for faulty methods. First, adjustment for unknown or unmeasured confounders is impossible. Second, adjustment for one confounder may create imbalance for another, since confounders are rarely fully correlated. For example, adjustment for age in the Swedish trials might seem reasonable; however in the New York trial, age was evenly distributed whereas several other prognostic factors were not.[16,21] Which adjustments should then be preferred for that trial? There must have been many other imbalances in prognostic factors at baseline in the Swedish trials, and there is a strong probability that other adjustments would have produced other results, both more and less extreme than a relative risk of 1·05 for the increase in total mortality with screening. Thus, the third important problem with adjustments is the risk of biased analyses when results of trials which were meant to be randomised but were found not to be so are adjusted post hoc.

The credibility of the Swedish meta-analysis is greatly weakened because it did not report that there were important imbalances at baseline in four of the five trials; that there was increased mortality in the screened groups; and that an adjustment for age had been made without being described.[10] The last point is particularly important, since readers would not have expected any adjustment to have been made in a meta-analysis of hundreds of thousands of women in which adjustments would not change anything, provided that the trials had been properly randomised. Shortly after the publication of the meta-analysis, Skrabanek obtained the mortality rates from the primary author and drew attention to the increased mortality in the screened groups[36] (10·0% *vs* 9·4%; relative risk 1·06). In their response,[37] Nyström and Larsson did not mention the imbalance in age, but defended the relative risk of 1·00 reported in the meta-analysis by comparing the observed number of deaths in the screened groups with the expected number in the population (15 695 *vs* 15 710). They also noted that the relative risks for total mortality in the individual trials were 0·98, 0·98, 0·99, 1·00, and 1·00. It is quite impossible, however, to have such rates for the individual trials and then an increased mortality of 1·06 (as we calculated) for the pooled analysis. Swift[38] noted subsequently that "a more precise and apt comparison is that between the mortality rates in the exposed and control groups". In response to this indisputable fact Nyström and Larsson wrote that "we prefer (see our response to Skrabanek) standardised relative risks to crude relative risks".[39] This reply is remarkable since the whole idea of randomisation is to make unbiased analyses possible, but it was another 3 years before Nyström and colleagues admitted publicly that the analysis of total mortality had been adjusted for age.[35]

Another serious flaw in the mammography trials is the fact that the number of randomised women was inconsistently reported for four of the six trials with inadequate randomisation methods. This inconsistency is not only odd, but it also raises further doubts about the validity of these trials.

The two trials with adequate randomisation found no effect of screening on mortality from breast cancer, not even a tendency towards an effect. By contrast, the pooled effect of the six trials with inadequate randomisation was highly significant. There was no overlap of the CIs for these two effect estimates. This lack of overlap is remarkable. Such disparate effects of subgroups of similar trials in a meta-analysis are very rare, and a strong warning signal that something is wrong. The explanation in such cases is generally methodological. In fact, the difference between the two point estimates, 1·05 and 0·75, is in good agreement with the results from empirical, methodological research. Randomised trials with inadequate or undescribed allocation methods exaggerate the estimated intervention effect by 33–41%, on average.[14,15] The bias can be even larger in cohort studies. For example, a meta-analysis of cohort studies of hormone replacement therapy showed protection against coronary heart disease (relative risk 0·50 [0·43–0·56]),[40] which was not confirmed in a large randomised trial (0·99 [0·80–1·22]);[41] again, there was no overlap of the 95% CIs.

The Canadian trial has been subjected to a fair amount of criticism, probably because it had the most negative results of the eight trials. The criticism has been rebutted;[26] somewhat ironically, this trial seems to be the one that is by far the best documented. A persistent criticism has been that an effect would be difficult to find because the breasts of all women in the age-group 50–59 years were physically examined regularly. This criticism is unwarranted because mammography will identify many tumours that are too small to be detected on physical examination alone. Furthermore, any effect of physical examination is likely to be small. A study of 122 471 women found no effect of regular self-examination of the breast on breast-cancer mortality after 9 years of follow-up, even though twice as many of the intervention group consulted an oncologist.[42] In addition, Kerlikowske's meta-analysis found that the regular clinical examinations in the non-Swedish trials had no influence on the relative risk.[43] We also much doubt the importance of the fact that the Canadian trial was not community based. Proper randomisation ensures the internal validity of a trial, and if mammography were effective, an effect should also be seen in a selected part of the population. Finally, the quality of the mammography has been criticised as being poor,[26] but the tumours found in the Canadian trial were smaller, on average, than those found in the Swedish trials.[44]

The study reports provided very few data on morbidity associated with screening. Some might argue that an increased occurrence of surgery, chemotherapy, and radiotherapy in the screened group is only natural and that, in the long run, over decades, the interventions would become less drastic because the tumours would be detected earlier. However, another point of view is that screening would be expected to increase morbidity in the

long run because of false-positive findings, cell changes that may never develop into cancer, and cancers that will develop so slowly that the woman dies of other causes before the cancer becomes apparent.

We could not assess psychological morbidity related to false-positive findings because this feature was not reported in the trials. In the USA, Elmore and colleagues[45] estimated that 49% of screened women will experience at least one false-positive mammogram during ten screening rounds and that 19% will be subjected to biopsy.[45] In the Swedish trials, false-positive rates of 4–6% have been reported,[9,28,29,31] corresponding to an average risk of 40% of a false-positive mammogram during ten rounds.

We conclude that screening for breast cancer with mammography is unjustified.

On the one hand, those who believe that the Swedish trials are unbiased have to accept from the data that screening for breast cancer with mammography causes more deaths than it saves. The total mortality in the five Swedish trials was 10%,[10] the relative risk of death was 1·06, and the Swedish meta-analysis showed a difference in breast-cancer mortality of 0·1% after 12 years of follow-up.[10] The data therefore show that for every 1000 women screened throughout 12 years, one breast-cancer death is avoided but the total number of deaths is increased by six.

On the other hand, those who believe the Swedish trials (apart from the Malmö trial) are biased have to accept that there is no reliable evidence that screening decreases breast-cancer mortality.

There is a need for further follow-up of the two unbiased trials and for detailed scrutiny of the other trials to see whether subgroups of women can be identified who have been properly randomised.

Contributors

Peter C Gøtzsche did the data searches and most of the analyses and wrote the drafts of the paper. Both researchers read the key articles critically and Ole Olsen contributed importantly to the final article.

Acknowledgments

The study was funded by the Danish Institute for Health Technology Assessment. We thank the following investigators for additional information on their trials: Samuel Shapiro, Ingvar Andersson, László Tabár, and Nils Bjurstam.

References

1 Dickersin K. Breast screening in women aged 40–49 years: what next? *Lancet* 1999; **353:** 1896–97.

2 Chu KC, Smart CR, Tarone RE. Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan clinical trial. *J Natl Cancer Inst* 1988; **80:** 1125–32.

3 Alexander FE, Anderson TJ, Brown HK, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* 1999; **353:** 1903–08.

4 Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1–breast cancer detection and death rates among women aged 40–49 years. *Can Med Assoc J* 1992; **147:** 1459–76.

5 Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 2–breast cancer detection and death rates among women aged 50 to 59 years. *Can Med Assoc J* 1992; **147:** 1477–88.

6 Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *BMJ* 1988; **297:** 943–48.

7 Tabar L, Fagerberg G, Chen HH, et al. Efficacy of breast cancer screening by age: new results from the Swedish Two-county Trial. *Cancer* 1995; **75:** 2507–17.

8 Frisell J, Lidbrink E, Hellstrom L, Rutqvist LE. Follow-up after 11 years: update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res Treat* 1997; **45:** 263–70.

9 Bjurstam N, Bjorneld L, Duffy SW, et al. The Gothenburg breast screening trial: first results on mortality, incidence, and mode of detection for women ages 39–49 years at randomization. *Cancer* 1997; **80:** 2091–99.

10 Nystrom L, Rutqvist LE, Walls S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993; **341:** 973–78.

11 Sjönell G, Ståhle L, Hålsokontroller med mammografi minskar inte dödlighet i bröstcancer. *Läkartidningen* 1999; **96:** 904–13.

12 Rehnqvist N, Rosén M, Karlberg I. Analys av dödligheten kräver helt annan metodik. *Läkartidningen* 1999; **96:** 1050–51.

13 Rutqvist LE. Naturalförloppet, grova metoder ledde till felkalkyl om böstcancer. *Läkartidningen* 1999; **96:** 1210–11.

14 Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273:** 408–12.

15 Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; **352:** 609–13.

16 Shapiro S, Venet W, Strax P, Venet L, eds. Periodic screening for breast cancer. Baltimore: Johns Hopkins University Press, 1998.

17 Shapiro S, Strax P, Venet L. Evaluation of periodic breast cancer screening with mammography: methodology and early observations. *JAMA* 1966; **195:** 731–38.

18 Fink R, Shapiro S, Roester R. Impact of efforts to increase participation in repetitive screenings for early breast cancer detection. *Am J Publ Health* 1972; **62:** 328–36.

19 Aron JL, Prorok PC. An analysis of the mortality effect in a breast cancer screening study. *Int J Epidemiol* 1986; **15:** 36–43

20 Shapiro S. Screening: assessment of current studies. *Cancer* 1994; **74:** 231–38.

21 Shapiro S. Evidence on screening for breast cancer from a randomized trial. *Cancer* 1977; **39:** 2772–82.

22 Alexander FE, Anderson TJ, Brown HK, et al. The Edinburgh randomised trial of breast cancer screening: results after 10 years of follow-up. *Br J Cancer* 1994; **70:** 542–48.

23 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.

24 Miller AB, Howe GR, Wall C. The National Study of Breast Cancer Screening Protocol for a Canadian randomized controlled trial of screening for breast cancer in women. *Clin Invest Med* 1981; **4:** 227–58.

25 Bailar IJC, MacMahon B, Phillips RA. Randomization in the Canadian National Breast Screening Study; a review for evidence of subversion. *Can Med Assoc J* 1997; **156:** 193–99.

26 Baines CJ. The Canadian National Breast Screening Study: a perspective on criticisms. *Ann Intern Med* 1994; **120:** 326–34.

27 Miller AB, To T, Baines CJ. Wall C. The Canadian national breast screening study: update on breast cancer mortality. *J Natl Cancer Inst Monogr* 1997; **22:** 37–41.

28 Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmö mammographic screening program. *J Natl Cancer Inst Monogr* 1997; **22:** 63–67.

29 Frisell J, Glas U, Hellstrom L, Somell A. Randomized mammographic screening or breast cancer in Stockholm: design, first round results and comparisons. *Breast Cancer Res Treat* 1986; **8:** 45–54.

30 Frinsell J, Eklung G, Hellstrom L. Lidbrink E, Rutqvist LE, Sonnell A. Randomised study of mammography screening-preliminary report on mortality in the Stockholm trial. *Breast Cancer Res Treat* 1991; **18:** 49–56.

31 Tabar L, Gad A, Holmberg L. Reduction in advanced breast cancer: results of the first seven years of mammography screening in Kopparberg, Sweden. *Diagn Imaging Clin Med* 1984 **54:**158–64.

32 Tabar L, Fagerberg CJ, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography: randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet* 1985; i: 829–32.

33 Fagerberg G, Baldetorp L, Gontoft O, Lundstrom B, Manson JC, Nordenskjold B. Effects of repeated mammographic screening on breast cancer stage distribution: results from a randomised study of 92 934 women in a Swedish county. *Acta Radiol Oncol* 1985; **24:** 465–73.

34 Tabar L, Fagerberg G, Duffy SW, Day NE. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health* 1989; **43:** 107–14.

35 Nystrom L, Larsson LG, Wall S, et al. An overview of the Swedish randomised mammography trials: total mortality pattern and the representivity of the study cohorts. *J Med Screening* 1996; **3:** 85–87.

36 Skrabanek P. Breast cancer screening with mammography. *Lancet* 1993; **341:** 1531.

37 Nyström L, Larsson L-G. Breast cancer screening with mammography. *Lancet* 1993; **341:** 1531–32.

38 Swift M. Screening mammography. *Lancet* 1993; **342:** 549–50.

39 Nyström L, Larsson L-G. Screening mammography. *Lancet* 1993; **342:** 550–51.

40 Stampfer MJ, Colditz GA. Estrogen replacement therapy and

coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med* 1991; **20:** 47–63.
41 Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women: Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* 1998; **280:** 605–13.
42 Semiglazov VF, Moiseenko VM, Protsenko SA, et al. Promezhutochnye rezul'taty programmy Rocciia (Santkt-Peterburg)/VOZ po otsenke effektivnosti samoobsledovaniia molochnykh zhelez (Preliminary results of the Russia (St Petersburg)/WHO program for the evaluation of the effectiveness of breast self-examination.) *Vopr Onkol* 1996; **42:** 49–55.
43 Kerlikowske K. Efficacy of screening mammography: a meta-analysis. *JAMA* 1995; **273:** 149–54.
44 Narod SA. On being the right size: reappraisal of mammography trials in Canada and Sweden. *Lancet* 1997; **349:** 1869.
45 Elmore JG, Barton MB, Moceri VM, et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998; **338:** 1089–96.

Review

Herb-drug interactions

Adriane Fugh-Berman

Concurrent use of herbs may mimic, magnify, or oppose the effect of drugs. Plausible cases of herb-drug interactions include: bleeding when warfarin is combined with ginkgo (*Ginkgo biloba*), garlic (*Allium sativum*), dong quai (*Angelica sinensis*), or danshen (*Salvia miltiorrhiza*); mild serotonin syndrome in patients who mix St John's wort (*Hypericum perforatum*) with serotonin-reuptake inhibitors; decreased bioavailability of digoxin, theophylline, cyclosporin, and phenprocoumon when these drugs are combined with St John's wort; induction of mania in depressed patients who mix antidepressants and *Panax ginseng*; exacerbation of extrapyramidal effects with neuroleptic drugs and betel nut (*Areca catechu*); increased risk of hypertension when tricyclic antidepressants are combined with yohimbine (*Pausinystalia yohimbe*); potentiation of oral and topical corticosteroids by liquorice (*Glycyrrhiza glabra*); decreased blood concentrations of prednisolone when taken with the Chinese herbal product xaio chai hu tang (sho-saiko-to); and decreased concentrations of phenytoin when combined with the Ayurvedic syrup shankhapushpi. Anthranoid-containing plants (including senna [*Cassia senna*] and cascara [*Rhamnus purshiana*]) and soluble fibres (including guar gum and psyllium) can decrease the absorption of drugs. Many reports of herb-drug interactions are sketchy and lack laboratory analysis of suspect preparations. Health-care practitioners should caution patients against mixing herbs and pharmaceutical drugs.

> "Poisons and medicines are oftentimes the same substances given with different intents."
>
> Peter Mere Latham (1789–1875)

Many medicinal herbs and pharmaceutical drugs are therapeutic at one dose and toxic at another. Interactions between herbs and drugs may increase or decrease the pharmacological or toxicological effects of either component. Synergistic therapeutic effects may complicate the dosing of long-term medications—eg, herbs traditionally used to decrease glucose concentrations in diabetes[1] could theoretically precipitate hypoglycaemia if taken in combination with conventional drugs.

Herbal medicines are ubiquitous: the dearth of reports of adverse events and interactions probably reflects a combination of under-reporting and the benign nature of most herbs used. Experimental data in the field of herb-drug interactions are limited, case reports scarce, and case series rare. This lack of data is also true of drug-drug interactions: published clinical studies are mainly case reports (controlled trials are scarce, since the random assignment of patients to trials that examine unintended effects is not ethical). The true prevalence of drug interactions is substantial but unknown. One study of 1000 elderly people admitted to a hospital from the emergency department found that 538 patients were exposed to 1087 drug-drug interactions; 30 patients experienced adverse effects as a consequence of these interactions.[2] In clinical practice, polypharmacy is common, and to the mixture physicians prescribe, patients add various over-the-counter medications, vitamins, herbs, and foods. All ingested substances have the potential to interact.

Lancet 2000; **355:** 134–38

George Washington University School of Medicine and Health Sciences, Department of Health Care Sciences, 2150 Pennsylvania Avenue, NW 2B-417, Washington, DC 20037, USA
(A Fugh-Berman MD)
(e-mail: fughberman@aol.com)

Source and extent of review

Sources for this review include MEDLINE 1966–98 (searched under MeSH terms "drug interactions" combined with "herbal medicine", "traditional medicine", "Chinese traditional medicine", "African traditional medicine", "Ayurvedic medicine", "Oriental traditional medicine", "Unani medicine", and "Arabic medicine"); EMBASE 1994–99 (searched under the same terms); reference dredging; and my own files on the subject.

Many reports of herb-induced interactions lack crucial documentation on temporal relations and concomitant drug use. Perhaps the most serious problem encountered in analysing such reports is the consistent absence of any effort (beyond that of reading the label) to establish a positive identification of the herb involved, and to exclude the effect of contaminants or adulterants. Unless noted otherwise, the reports mentioned herein did not include chemical analyses.

This review was limited to the most commonly used medicinal plants, and to clinical reports (animal studies

COMMENTARY

Assessment of nationwide cancer-screening programmes

See page 129

Screening for cancer has always been highly controversial, partly because the procedure is for seemingly healthy people, for whom the benefit should be clear cut. Evidence of this benefit is, however, for the group as a whole. At the individual level, prediction of who will benefit and who will suffer more harm than good is impossible. The balance between favourable and unfavourable effects is delicate.[1]

The best evidence is that provided by randomised trials. Breast-cancer screening is perhaps one of the most intensively evaluated health-care practices, with eight completed randomised trials, in which half a million women have taken part. Still, trial results will not necessarily be replicated in screening programmes since "Trial results represent the results of the enthusiasts, the pioneers in this field who with evangelical style have developed screening and by observational attention to minor abnormalities on mammograms have reduced the mortality from breast cancer".[2]

In today's *Lancet*, 40 years after the start of the first breast-screening trial, Peter Gøtzsche and Ole Olsen challenge whether this mortality reduction exists. The report that the expected reductions in breast cancer mortality rate had not materialised in Sweden, where five trials had been instrumental in encouraging a nationwide screening programme, had prompted Gøtzsche and Olsen to examine the randomisation process of trials run in the 1960s to 1980s. In only two trials, one in Malmö and the other in Canada,[5] was the age-distribution study and control groups perfectly matched. These two methodologically adequate trials overall did not show a decrease in mortality from breast cancer. Gøtzsche and Olsen thus reasoned that, since trials that did show a reduction were biased, breast-cancer screening does not reduce disease-related mortality and is therefore unjustified.

What is wrong with this reasoning? First, the conclusion that all the trials other than the Malmö and Canadian trials were not adequately randomised should be examined. Both the Edinburgh trial[6] and Kopparberg/Östergötland trials[7,8] were randomly selected from each block for participation in the screening or control groups. The women in the parishes and municipalities in the Kopparberg/Östergötland trials could have differed more than those in the areas randomised to the control group. Is this difference crucial and does it considerably weaken credibility? If anything, any bias produced in the Kopparberg/Östergötland trials (which was the trial that showed the largest benefits for women aged 50 and over) should mainly counteract the reduction of breast-cancer mortality since the age distribution could have increased the detection of cancer in study groups.

Gøtzsche and Olsen, with their experience with meta-analyses for the Cochrane Collaboration, show that the small age differences and different randomisation procedures provide explain why there was no change in confidence intervals between the results of the trials that showed no mortality reduction (the Malmö and the Canadian trials) and those of the trials that did (Goteborg, Stockholm, and Kopparberg/Östergötland). However, in screening trials enrolling tens of thousands of women, very small age differences between groups become statistically significant. A statistically significant difference in age is not necessarily clinically significant or relevant when in screening trials. It is therefore unjustified to base judgment of whether a trial is biased on "statistically significant" differences. Breast-screening trials have made much effort to prevent serious biases in design[9]—eg, use of independent assessors to evaluate cause of death, by making assessors unaware of group allocation, study use of different end-points for follow-up. Other variables, such as screening interval, age-groups, contamination rate, and duration of follow-up also have tremendous impact on outcomes of screening trials.[9] A standard meta-analysis does not take into account the variables that contribute to relative risks for specific screening situations. Screening trials should not be treated in the same way as a standard therapeutic trial. In therapeutic trials participants have the disease, the main variables are treatment or no treatment and doses of treatment, study populations are much smaller than those in screening trials, and age differences are more crucial than in screening trials. In screening, the attendance rate, quality of the mammograms, accuracy in reading of the films, decisions for referral, and distribution of clinical stage before the start of the programme are all crucial factors influencing effect of screening. The quality of the mammograms in the Canadian trial is on record as having been poor at the start of the trial.[10] The control population in Canada seemed naturally to have a relatively favourable distribution of cancer.[11] The Malmö trial had the lowest

attendance rates and may well have had one of the largest contamination rates in the control group. The most recent results of the Malmö trial[12] show a relative risk of death from breast cancer of 0·81 (ages 45–69), not mentioned by Gøtzsche and Olsen, and a 26% reduction in breast-cancer mortality for women aged 55–69 years. Obviously, these authors have focused on one particular (and certainly very important) issue, but have disregarded the fact that other factors probably have a more important part in lowering the mortality rate through screening. In fact, Gøtzsche and Olsen seem to regard the use of a different randomisation process as being synonymous with a tendency towards a lower quality in execution and analysis of the study. Anyone who has visited the Swedish trial centres and looked at their mammograms will find this association unlikely to be the case.

What can be learnt from Gøtzsche and Olsen's report? The designing, running, and evaluation of cancer-screening trials and programmes is a huge task, and their assessment and the presentation of the data have to be as precise as possible, with explicit accounts of reasons for possible changes in numbers, design, or analysis (from the one report to the other). Furthermore, there should not be any lack of clarity about the exact randomisation process.

Evaluation of the outcome of cancer screening at a national level is very much a long-term proposition. In the UK and the Netherlands, breast cancer in women aged 50 and over is being detected earlier than in the past because of screening.[13,14] Moreover, laboratory work has shown that the treatment of small cancers before a critical number of blood vessels has formed (as tumours enlarge or become poorly differentiated) may prevent metastasis.[15] In the UK, there has been a clear reduction in breast-cancer mortality, due in part to the national breast-screening programme.[16,17] The Finnish programme, which has been built up cohort by cohort, seems to indicate likewise.[18]

Breast-cancer mortality in the Netherlands for women aged 60–69 is falling.[17] This screening programme was introduced in 1989 and recruited women gradually until by 1997 all eligible women had been invited at least once. Today some 800 000 women are screened per year. No statistically significant breast-cancer mortality reduction has been expected, or found, in the first 9 years of the programme. Is the lack of an effect on breast-cancer mortality in Sweden thus surprising? No conclusion can be drawn without clear information about the screening process and its quality across Sweden. If no other factors are involved, lead time, mean survival, and build-up period all result in a lag time.

All efforts should be put into evaluation at the individual level of mortality statistics, linking the causes of death in women to screening exposure in the nationwide programmes. However, the evidence from current programmes will never be as strong as that from randomised trials. Publication of the reports, whether the results are positive or negative, together with the seeking of the opinions of women, are the ways by which an answer can be reached as to whether or not screening programmes are justified and at what cost to women and to society.

I am a member of the National Evaluation Team for breast-cancer screening in the Netherlands.

Harry J de Koning
Department of Public Health, Erasmus University Rotterdam, 3000 DR Rotterdam, Netherlands

1 de Haes JCJM, de Koning HJ, van Oortmarssen GJ, van Agt HME, de Bruyn AE, van der Maas PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. *Int J Cancer* 1991; **49:** 538–44.
2 Fentiman I. Detection and treatment of early breast cancer. Philadelphia: Lippincott, 1990: 618.
3 Sjönell G, Ståhle L. Mammographic screening does not reduce breast cancer mortality. [in Swedish] *Läkartidningen* 1999; **96:** 904–13.
4 Andersson I, Aspegren K, Janzon L et al. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *BMJ* 1988; **297:** 943–48.
5 Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *CMAJ* 1992; **147:** 1477–88.
6 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.
7 Tabár L, Gad A. Screening for breast cancer: the Swedish trial. *Radiology* 1981; **138:** 219–22.
8 Fagerberg G, Baldetorp L, Gröntoft O, Lundström B, Månson JC, Nordenskjöld B. Effects of repeated mammographic screening on breast cancer stage distribution. *Acta Radiologica Oncol* 1985; **24:** 465–73.
9 de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PMM, van der Maas PJ. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. *J Natl Cancer Inst* 1995; **87:** 1217–23.
10 Baines CJ, Miller AB, Kopans DB, et al. Canadian National Breast Screening Study: assessment of technical quality by external review. *AJR Am J Roentgenol* 1990; **155:** 743–47.
11 Narod SA. On being the right size: a reappraisal of mammography trials in Canada and Sweden. *Lancet* 1997; **349:** 1846.
12 Andersson I, Nyström L. Mammography screening. *J Natl Cancer Inst* 1995; **87:** 1263–64.
13 Moss SM, Michel M, Patnick J, Johns L, Blanks R, Chamberlain J. Results from the NHS breast screening programme 1990–1993. *J Med Screen* 1995; **2:** 186–90.
14 Fracheboud J, de Koning HJ, Beemsterboer PMM, et al. Nation-wide breast cancer screening in The Netherlands: results of initial and subsequent screening 1990–95. *Int J Cancer* 1998; **75:** 694–98.
15 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:** 1120–24.
16 Quinn M, Allen E. Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. *BMJ* 1995; **311:** 1391–95.
17 van den Akker-van Marle E, de Koning H, Boer R, van der Maas P. Reduction in breast cancer mortality due to the introduction of mass screening in the Netherlands: comparison with the United Kingdom. *J Med Screen* 1999; **6:** 30–34.
18 Hakama M, Pukkala E, Heikkila M, Kallio M. Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *BMJ* 1997; **314:** 864–67.

Maternal blood pressure and birthweight

See page 87

Hypertension in pregnancy is a leading cause of maternal and neonatal mortality and morbidity, so a large proportion of antenatal care goes to the detection and management of this disorder. Antihypertensive drug therapy is an important part of management strategy. The effect of such therapy on the baby's birthweight has been examined by P von Dadelszen and colleagues in their meta-analyses, reported in today's *Lancet*, of up to 45 randomised controlled trials.

The investigators noted the effect of change in mean arterial blood pressure (MAP) in a novel way. MAPs were calculated for blood pressure at entry to the trial and before delivery. Changes in MAPs (ΔMAP) between these two times were compared between drug and placebo groups. A ΔMAP of 10 mm Hg meant that the MAP had fallen by 10 mm Hg more in the treatment than in the placebo group. Drug/drug and drug/placebo comparisons were made. For drug/drug comparisons β-blockers were arbitrarily assigned to be the experimental intervention and methyldopa to be the control. 38 trials related to therapy