

Relation between breast cancer mortality and screening effectiveness: systematic review of the mammography trials

Peter C. Gøtzsche

ABSTRACT

The mammography screening trials have shown varying results. This could be because screening was better in some trials than in others at advancing the time of diagnosis. If so, more cancers would be identified in such trials relative to the control group, and fewer of the cancers would have reached an advanced stage. I performed a systematic review of the mammography screening trials using meta-regression. Finding many cancers was not related to the size of the reduction in breast cancer mortality ($p = 0.19$ after seven and $p = 0.73$ after 13 years of follow-up). In contrast, finding few cancers in stage II and above predicted a larger reduction in breast cancer mortality ($p = 0.04$ and $p = 0.006$). This expected association was also found for node-positive cancers ($p = 0.008$ and $p = 0.04$). However, a screening effectiveness of zero (same proportion of node-positive cancers in the screened group as in the control group) predicted a significant 16% reduction in breast cancer mortality after 13 years (95% confidence interval, 9% to 23% reduction). This can only occur if there is bias. Further analyses uncovered bias in both assessment of the cause of death and of the number of cancers in advanced stages. Consequently, the differences in the reported reductions in breast cancer mortality cannot be explained by differences in screening effectiveness. Given that the size of the bias was similar to the estimated screening effect, screening appeared ineffective.

The randomised mammography screening trials have shown varying results. After 13 years of follow-up, the results range from a 42% decrease to a 2% increase in breast cancer mortality [1]. Debates about how these differences are best explained have mainly focused on trial quality, as some trials appear to be more reliable than others [1-3]. The most straightforward explanation – differences in screening effectiveness – has received little attention. Screening effectiveness can be perceived as the ability to advance the time of diagnosis, which leads to identification of more cancers than in an un-screened control group [3]. A screening programme that finds many cancers, e.g. owing to a high sensitivity, should therefore lead to a larger reduction in breast

cancer mortality relative to a control group than a programme that identifies fewer cancers.

One would also expect trials that were more effective in identifying cancers before they had metastasised to yield larger effects [3]. An indication that this may be the case was provided in a Letter to the Editor in *The Lancet* [4]. The authors found an association between the risk ratio for detecting node-positive cancers and the risk ratio for breast cancer mortality [4], but they included only women in the age-group 40-49 years and did not describe their methods.

The objective of this systematic review of the randomised mammography screening trials was to examine whether there is a relation between screening effectiveness and breast cancer mortality.

MATERIAL AND METHODS

The primary analysis was a linear regression (meta-regression) analysis weighted by the inverse variance for breast cancer mortality in the trials. This analysis related the screening effectiveness, defined as the log risk ratio (RR) of being diagnosed with cancer (including carcinoma in situ) within the first seven years to the log RR of breast cancer mortality after seven and 13 years, respectively, as the outcome.

In additional regression analyses, the RRs of stage II+ cancers (those that are either node-positive or at least 2 cm in size) and of node-positive cancers were used as explanatory variables.

Comprehensive Meta Analysis version 2.2.030, July 2006, was used (random effects model, unrestricted maximum likelihood).

Searches

The literature search was extensive. I searched PubMed with (breast neoplasms [MeSH] OR "breast cancer" OR mammography [MeSH] OR mammograph*) AND (mass screening [MeSH] OR screen*) and combined this search with a search on author names [1]. The latest search was performed in November 2008, and 24,479 records were imported into ProCite and searched for author names, cities and trial eponyms. Reference lists were scanned

REVIEW ARTICLE

The Nordic Cochrane Centre, Rigshospitalet and University of Copenhagen, Department 3343

Dan Med Bul
2011;58(3):A4246

and letters, abstracts, grey literature and unpublished data were included.

A total of nine trials were found. They were performed in New York, Canada (two trials), the UK (two trials) and Sweden (four trials: Two-County (sometimes reported separately for the two counties, Kopparberg and Östergötland), Malmö, Stockholm and Göteborg (divided in two sub-trials by age)). The age range 45-64 years was covered by most trials [1], but the UK Age Trial only included women between 39 and 41 years of age [5].

Data

Trial data on relative risks for breast cancer mortality after seven and 13 years from our 2009 Cochrane review were used [1]. Furthermore, I extracted data from the many papers included in this review on total number of cancers (including carcinoma in situ) and number of advanced cancers (number in stages II-IV and number that were node-positive).

Data on breast cancers from the majority of the trials vary from publication to publication, mostly because of changing cut-points for registration, different age groups and varying numbers of women in the analyses [1]. All the retrieved data were entered into an

Excel spreadsheet and extensive validity checks were performed, e.g. calculation of relative risks for finding cancers and cancers in specific stages and comparison of the results. Data used in the statistical analyses were checked again by comparing them with trial report data. In some cases, the data on the total number of cancers and the number of women (the denominators for the calculations) were slightly different from those of the Cochrane review [1], as data divided on stage and node-positivity were used in the present review. However, differences were immaterial, as the RRs for cancer detection were either identical or very similar to those of the review, the largest difference being 0.05 (1.44 rather than 1.49 in the Stockholm trial).

Data were available from all trials on breast cancer mortality and on total number of cancers: Canada [6-10], Malmö [11, 12], Kopparberg [13-15], Östergötland [13, 14], Stockholm [12, 16, 17], Göteborg [18-21], New York [22, 23], Edinburgh [24, 25] and the UK Age trial [5, 26]. Other papers provided additional information on the type of cancers [27-33].

Specific issues in the individual trials

In New York, about the same number of cancers was detected in the screened group and the control group, and it is therefore surprising that a large effect was reported [1]. However, the cause-of-death assessment seems to have been biased, and some cancers in the control group – and their associated deaths – should have been excluded, as these patients were diagnosed with breast cancer prior to randomisation [1]. The Edinburgh trial was cluster-randomised, but this worked so poorly that 26% of the women in the control group and 53% in the study group belonged to the highest socioeconomic level. This resulted in mammographic screening being associated with a 26% reduction in cardiovascular mortality among invited women [1], a result that cannot have been caused by screening. Sensitivity analyses were therefore performed that excluded the data from the two trials.

Apart from the Malmö trial, the Swedish trials screened the whole control group 3-5 years after randomisation [1]. Therefore, the number of cancers found *before* the control group screen was used to avoid this serious contamination. In additional analyses, however, the contamination was disregarded and the additional cancers found at the control screening were included.

In Göteborg, the number of cancers detected before the control group was screened was only available for the youngest age group, 39-49 years [20], whereas number of deaths after seven years was only available for the slightly narrower age group 40-49 years [18]. Varying denominators have also been reported for the other Swedish trials, and the denominators that cor-



APPENDIX

Regression analysis

Result from the regression analysis of node-positive cancers as a predictor of the reduction in breast cancer mortality after 13 years, as shown in Figure 2d.

Slope 0.451, SE 0.224, $p = 0.045$.

Intercept -0.175 , SE 0.043, $p = 0.00005$.

Thus, $\ln RR_{\text{effect}} = 0.451 \times \ln RR_{\text{node-pos}} - 0.175$

For $RR_{\text{node-pos}} = 1$, $RR_{\text{effect}} = e^{-0.175} = 0.84$, i.e. a 16% reduction in breast cancer mortality.

The 95% confidence interval for this estimate is $e^{-0.175 \pm 2 \times 0.043} = 0.77$ to 0.91 , i.e. a 9% to 23% reduction in breast cancer mortality.

$\ln RR_{\text{effect}} = \ln$ risk ratio for breast cancer mortality

$\ln RR_{\text{node-pos}} = \ln$ risk ratio for node-positive cancers

SE = standard error

Unrecognized node-positive cancers in the control group

In the Canadian trial, which included women aged 40-49 years, there were 23 breast cancer deaths among women with node-negative cancers in the study group and 34 breast cancer deaths in the control group [37], which yields a risk ratio of $23/34 = 0.68$ for breast cancer deaths among women with node-negative cancers.

However, the risk ratio for breast cancer death for the whole trial was 1.14 at that particular point in time, after 10.5 years of follow-up [37]. Assuming that this risk ratio applies to both node-negative and node-positive cancers, which seems reasonable as so many women with node-negative cancer died, there were far too many deaths among women with node-negative cancers in the control group. We can calculate the expected number of deaths, x , from $23/x = 1.14$, which gives $x = 20.2$. Thus, there were 14 deaths too many.

This suggests that 14 cancers that were labelled node-negative in the control group were actually node-positive.

The reported number of node-positive cancers in the control group was 66 [6]. If we add the 14 cancers to this number, we get 80 in total, or 1.21 as many (80/66) as those reported.

responded to the number of deaths may therefore be slightly different from those that corresponded to the number of cancers.

In Östergötland, there were no data on the number of cancers that included the control group screen after about seven years, but data existed after a more extended follow-up period. The number of cancers in the study group had increased by only 16% after this additional follow-up [14]. These data were used in the analyses, as only the RRs of cancers were needed, and these ratios differed very little in the trials when the total number of cancers differed as little as was the case in Östergötland.

Tumour data from the control group in Stockholm

had been multiplied by a factor that corresponded to the smaller size of this group compared with the screened group [17]. The data were re-corrected for the analyses by dividing with this factor.

Data on breast cancer mortality and on the number of cancers are shown in **Table 1** and **Table 2**, respectively.

RESULTS

Screening effectiveness

measured as total number of cancers

Screening effectiveness, defined as the RR for the total number of detected cancers, was not related to the reduction in breast cancer mortality, $p = 0.19$ after seven



TABLE 1

Data on breast cancer mortality.

Study and age group (yrs)	Number of women				Number of breast cancer deaths				RR for breast cancer mortality	
	SG, 7 yrs	CG, 7 yrs	SG, 13 yrs	CG, 13 yrs	SG, 7 yrs	CG, 7 yrs	SG, 13 yrs	CG, 13 yrs	7 yrs	13 yrs
Ca1, Canada (40-49)	25,214	25,216	25,214	25,216	38	28	105	108	1.36	0.97
Ca2, Canada (50-59)	19,711	19,694	19,711	19,694	38	39	107	105	0.97	1.02
Ma, Malmö (45-70)	21,088	21,195	20,695	20,783	63	66	87	108	0.96	0.81
Ko, Kopparberg (40-74)	39,051	18,846	38,589	18,582	71	52	126	104	0.66	0.58
Ös, Östergötland (40-74)	39,034	37,936	38,491	37,403	53	67	135	173	0.77	0.76
TC, Two-County (40-74)	78,085	56,782	77,080	55,985	124	119	261	277	0.76	0.68
St, Stockholm (40-64)	38,525	20,651	40,318	19,943	53	40	66	45	0.71	0.73
G1, Göteborg (40-49)	10,821	13,101	11,724	14,217	6	10	34	59	0.73	0.70
G2, Göteborg (50-59)	9,926	15,744	9,926	15,744	NA	NA	54	103	NA	0.83
NY, New York (40-64)	31,000	31,000	31,000	31,000	81	124	218	262	0.65	0.83
Ed, Edinburgh (45-64)	23,226	21,904	28,628	26,015	68	76	176	187	0.84	0.86
Age, UK Age Trial (39-41)	53,884	106,956	53,884	106,956	NA	NA	105	251	NA	0.83
Total	313,714	332,387	395,260	391,538	471	502	1,213	1,505		

CG = control group; NA = not available; RR = risk ratio; SG = study group.



TABLE 2

Data on breast cancers; values in [] include the control group screen.

Study and age group (years)	Number of cancers			Number of cancers in stage II+				Number of node-positive cancers			
	SG	CG	RR	SG	CG	SU	RR	SG	CG	SU	RR
Ca1, Canada (40-49)	426	327	1.30	143	120	5	1.19	102	66	87	1.55
Ca2, Canada (50-59)	460	365	1.26	141	142	5	0.99	98	90	101	1.09
Ma, Malmö (45-70)	588	447	1.32	190	231	13	0.83	NA	NA	NA	NA
Ko, Kopparberg (40-74)	694 [676] ^a	255 [359]	1.31 [0.91]	228 [NA]	151 [NA]	NA	0.73 [NA]	NA	NA	NA	NA
Ös, Östergötland (40-74)	621 [720]	464 [682]	1.30 [1.05]	181 [NA]	225 [NA]	NA	0.78 [NA]	NA	NA	NA	NA
TC, Two-County (40-74)	See Ko+Ös	See Ko+Ös	See Ko+Ös	See Ko+Ös	See Ko+Ös	See Ko+Ös	See Ko+Ös	325 [325]	268 [323]	275 [318]	0.88 [0.73]
St, Stockholm (40-64)	371 [428]	127 [217]	1.44 [0.98]	143 [173]	74 [104]	0	0.96 [0.82]	NA	NA	NA	NA
G1, Göteborg (40-49)	144 [144]	151 [195]	1.16 [0.90]	NA	NA	NA	NA	39 [NA]	73 [NA]	NA	0.65
G2, Göteborg (50-59)	NA [44]	NA [72]	NA [0.97]	NA	NA	NA	NA	46 [NA]	71 [NA]	NA	1.03
NY, New York (40-64)	426	439	0.97	162	190	31	0.86	102	121	59	0.84
Ed, Edinburgh (45-64)	395	268	1.39	228	221	6	0.97	145	144	227	0.95
Age, UK Age Trial (39-41)	482	821	1.17	NA	NA	NA	NA	124	276	248	0.89
Total	4,607	3,664									

CG = control group; NA = not available; RR = risk ratio; SG = study group; SU = stage unknown.

a) Fewer cancers than without control group screen, as these data come from another paper and data for this trial were not consistently reported.

years (Figure 1a) and $p = 0.73$ after 13 years of follow-up (Figure 1b). Figure 1a shows a clustering of widely varying mortality estimates for approximately the same screening effectiveness. Furthermore, the New York trial is an outlier that unduly influences the analysis, shifting the regression line upwards, although a downward trend is expected, as detection of more cancers in the screened group should decrease breast cancer mortality. Regression analysis after exclusion of the trials from New York and Edinburgh (see Material and methods) is more appropriate, but did not change the findings ($p = 0.43$ after seven years and $p = 0.61$ after 13 years).

When the cancers detected at the control group screen were included, there was a significant relationship between screening efficiency and the reduction in breast cancer mortality with good fits to the regression lines (Figures 1c and 1d). However, this relationship was the opposite of that which was expected. The more similar the number of cancers in the screening and the control groups, the larger the effect ($p = 0.02$, both after seven and 13 years). This relationship remained after

exclusion of the New York and Edinburgh trials ($p = 0.02$ and $p = 0.005$, respectively).

Screening effectiveness measured as advanced stage cancers

For cancers in stage II and above, a significant relationship in the expected direction was found, i.e. fewer advanced cancers in the screened group than in the control group predicted a larger reduction in breast cancer mortality, both after seven years ($p = 0.04$) and 13 years ($p = 0.006$) (Figures 2a and 2b). This relationship remained after exclusion of the New York and Edinburgh trials ($p = 0.04$ and $p = 0.006$, respectively).

Also for node-positive cancers, the expected trends were significant ($p = 0.008$ after seven years and $p = 0.04$ after 13 years) (Figures 2c and 2d). This relationship persisted also after exclusion of the New York and Edinburgh trials ($p = 0.03$ and $p = 0.02$, respectively).

Evidence of bias

The four regression lines for advanced cancers predicted a relative risk in breast cancer mortality ranging from

FIGURE 1

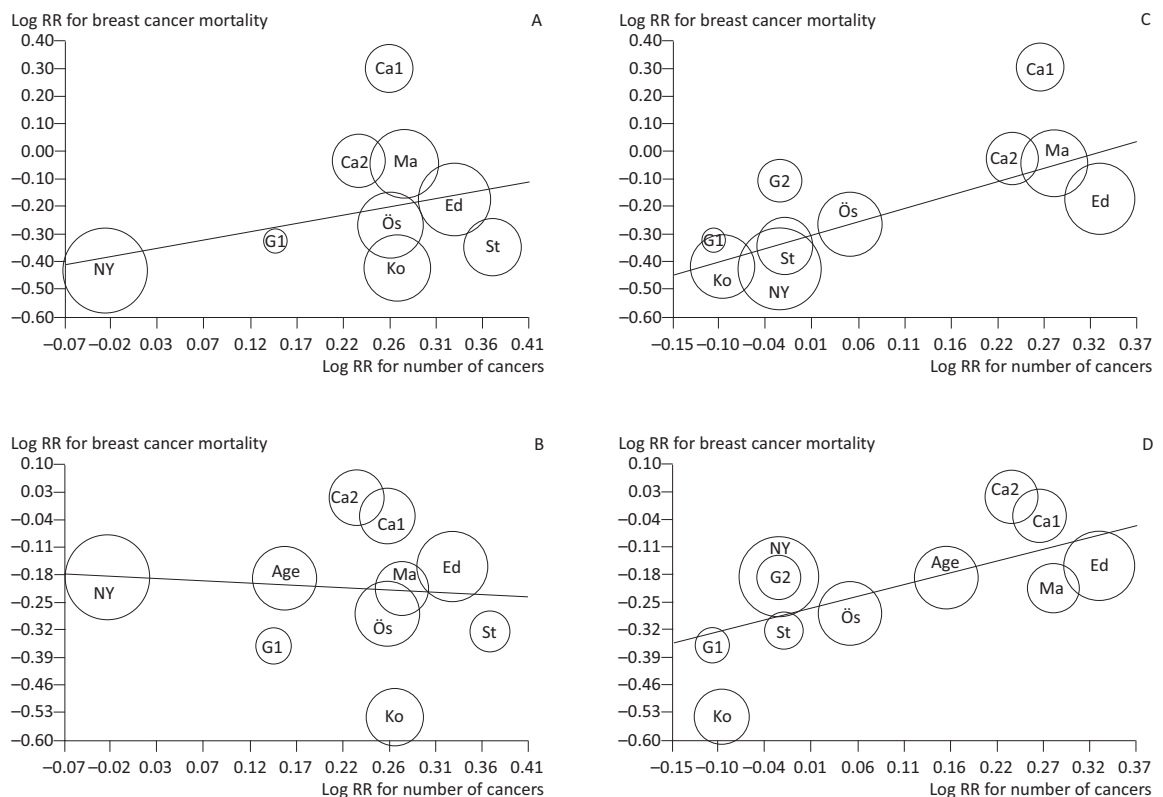
Meta-regressions of the risk ratio for detecting breast cancer and the risk ratio for dying from breast cancer. The circle areas are proportional in size to the weights. The x-axis shows screening effectiveness and the y-axis shows the effect of the screening.

A. Total number of cancers, breast cancer mortality after 7 years.

B. Total number of cancers, breast cancer mortality after 13 years.

C. Total number of cancers, incl. control group screen, breast cancer mortality after 7 years.

D. Total number of cancers, incl. control group screen, breast cancer mortality after 13 years.



RR = risk ratio

Age = UK Age Trial, 39-41 years; Ca1 = Canada, 40-49 years; Ca2 = Canada, 50-59 years; Ed = Edinburgh, 45-64 years; G1 = Göteborg, 40-49 years; G2 = Göteborg, 50-59 years; Ko = Kopparberg, 40-74 years; Ma = Malmö, 45-70 years; NY = New York, 40-64 years; St = Stockholm, 40-64 years; Ös = Östergötland, 40-74 years.

0.84 to 0.91 for zero screening effectiveness (same proportion of advanced cancers in the screened group as in the control group, i.e. $RR = 1$ for number of cancers, and $\log RR = 0$). In the most powerful analysis, which was after 13 years for node-positive cancers (Figure 2d), a screening effectiveness of zero predicted a relative risk of 0.84 for breast cancer mortality. This 16% reduction in breast cancer mortality was highly significant ($p < 0.001$; 95% confidence interval, 9% to 23% reduction, see appendix for details). This can only occur if there is bias, as it is not possible to obtain an effect with a screening effectiveness of zero.

DISCUSSION

Screening advances the time of diagnosis, and the total number of cancers detected in a screened group relative to the number detected in a control group is therefore an unbiased measure of screening effectiveness [3]. Some of the screening-detected cancers were not destined to cause symptoms or death in the women's remaining lifetime [34], but the extent of this overdiagnosis is rather closely related to the ability to advance the

time of diagnosis because when the lead-time is longer, more women will die from other causes before their cancers become symptomatic.

The better screening is at advancing the time of diagnosis, the more cancers will be found in a screened group compared with a control group. Furthermore, fewer of these cancers will be advanced, which is the objective of screening. Thus, an effective screening programme would be expected to yield a relatively *large* RR for the total number of cancers detected and a relatively *low* RR for the number of advanced cancers. It was therefore surprising that there was no relation between breast cancer mortality and screening effectiveness calculated on the basis of the *total number* of cancers, given that – in the same trials – breast cancer mortality was clearly more reduced in those trials that had *fewer advanced* cancers in the screened group.

This discrepancy and the fact that zero screening effectiveness was associated with a 16% reduction in breast cancer mortality suggest that the number of advanced cancers or the number of breast cancer deaths, or both, is biased in favour of screening. For simplicity,

FIGURE 2

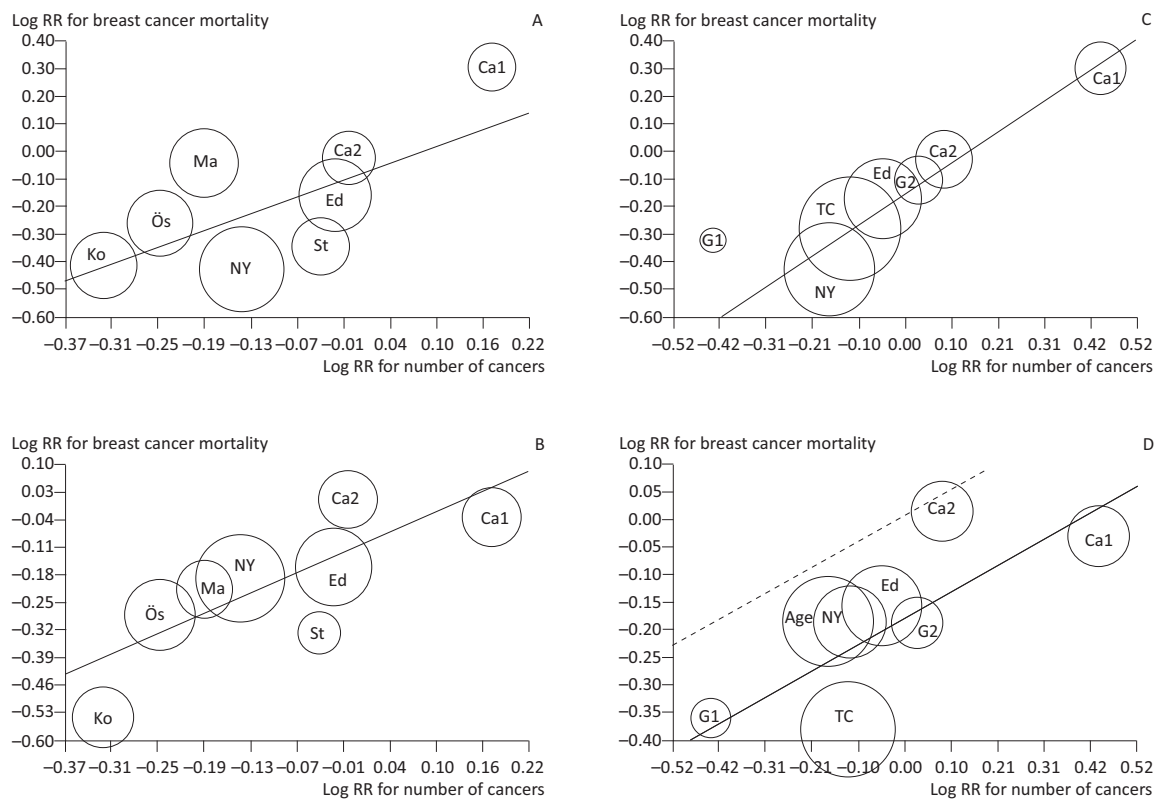
Meta-regressions of the risk ratio for detecting breast cancer and the risk ratio for dying from breast cancer. The circle areas are proportional in size to the weights. The x-axis shows screening effectiveness and the y-axis shows the effect of screening.

A. Number of cancers in stage II and above, breast cancer mortality after 7 years.

B. Number of cancers in stage II and above, breast cancer mortality after 13 years.

C. Number of node-positive cancers, breast cancer mortality after 7 years.

D. Number of node-positive cancers, breast cancer mortality after 13 years. The hatched line represents an unbiased regression line that crosses (0.0) and has the same slope as the calculated regression line.



RR = risk ratio

Age = UK Age Trial, 39-41 years; Ca1 = Canada, 40-49 years; Ca2 = Canada, 50-59 years; Ed = Edinburgh, 45-64 years; G1 = Göteborg, 40-49 years; G2 = Göteborg, 50-59 years; Ko = Kopparberg, 40-74 years; Ma = Malmö, 45-70 years; NY = New York, 40-64 years; St = Stockholm, 40-64 years; TC = Two-County, 40-74 years; Os = Östergötland, 40-74 years.

the influence of each potential bias will be explored separately below under the assumption that the other bias does not exist and using data on node-positive cancers and mortality after 13 years.

Bias in number of node-positive cancers

Many values were missing. The number of node-positive cases was only twice that of cases with unknown nodal status (Table 2). Node-negative cancers are not relevant, as many of these are overdiagnosed. Metastatic disease is considered the best proxy for breast cancer mortality, but more women with positive nodes failed to be identified in the control group than in the study group, as control group women were more likely to be treated in centres where careful nodal dissection was not the norm. This problem has been acknowledged for the Two-County and Canadian trials [35, 36] and is supported by the finding that in the Canadian trial covering women in the age group 40-49 years, 47% of those who died of breast cancer in the control group had node-negative cancer compared with only 28% in the mammography group [1, 37].

An estimate of the size of this bias can be obtained from the Canadian trial [6, 37]. Based on the RR for breast cancer mortality and assuming that this risk applies to both node-negative and node-positive cancers, which seems reasonable, as so many women with node-negative cancer died, there should be 1.2 times more node-positive cancers in the control group than actually reported (see appendix for details). If we multiply the number of node-positive cancers in the control groups of each trial in Table 2 by 1.2, the regression analysis shows a reduction in breast cancer mortality of 9% for zero screening effectiveness. Thus, if the Canadian findings can be generalized, about half of the 16% observed bias can be explained by underreporting of node-positive cancers in the control group.

Bias in assessment of cause of death

Assessment of the cause of death is inevitably biased in favour of screening, even when data from official cause-of-death registers are used [1]. One reason for this is that women who are screened are more likely to receive radiation treatment than controls, leading to an increased mortality from other causes and also to a reduction in local breast cancer recurrence. This makes it more likely that screened women with breast cancer will be assigned another cause of death [1].

The 16% bias in the regression analysis would disappear if we multiplied the number of breast cancer deaths in Table 1 in the screened group by 1.2, which would lead to zero effect for zero screening effectiveness. The factor 1.2 means that an additional 20% breast cancer deaths were missed in the screened groups. This

may seem unrealistic, but the Östergötland trial shows that it can occur. The Östergötland investigators, who were not blinded when they assessed the cause of death, reported a 24% reduction in breast cancer mortality, whereas the official cause-of-death register showed only a 10% reduction [1]. The difference between 24% and 10% corresponds to an additional 19% of breast cancer deaths in the screened group.

Data to facilitate an estimate of this bias are lacking from other trials, apart from the New York Health Insurance Plan (HIP) trial. In the New York HIP trial, differential misclassification may be responsible for about half of the reported breast cancer mortality reduction since a similar number of dubious cases were selected for blinded review from each group, while a much smaller proportion of the screened group was finally classified as having died from breast cancer [38].

Limitations

The assumption of linearity appears reasonable. Although the trials spanned almost 30 years, the data points for advanced cancers (Figure 2) were nicely distributed around the regression lines. Furthermore, the choice of statistical model was immaterial. A fixed effect model is usually not recommended for meta-regression, but it gave the same result for the most powerful analysis as the random effects model. I did not incorporate the variance in the number of cancers in the analyses, but that would not have made any material difference either. The greatest uncertainty stems from the mortality estimates because of the relatively small number of events.

The sensitivity and specificity of mammographic readings in the trials seem not to have changed since the New York trial [1]. It is therefore difficult to understand why the trials from Kopparberg, Östergötland, Stockholm and Göteborg, which screened the whole control group 3-5 years after randomisation and therefore had small intervention contrasts, were those that reported the largest reductions in breast cancer mortality after 13 years [1]. I included all trials, also the two flawed trials from New York and Edinburgh, to avoid accusations of selective reporting, and to facilitate a comparison with the smaller study of node-positive cancers in women aged 40 to 49 years [4], but it made no difference to the results whether or not these trials were included.

It could be argued that it is an oversimplification to suggest that a screening effectiveness of zero should lead to zero effect on breast cancer mortality. Screening brings forward the diagnosis of both localised and advanced cancers, and one might therefore theoretically see more advanced cases in a screened group than in a control group and still reduce breast cancer mortality. However, such a possibility is of minor relevance com-

pared with the biases identified in the present study, and, more importantly, it cannot explain them because most trials have *fewer* advanced cancers in the screened groups than in the control groups (see Figures 2a-d).

Implications for observational studies

The biases I identified were substantial. Furthermore, surgical and pathological expertise is likely to vary considerably between regions and over time. This suggests that comparative observational studies across regions, countries or time periods may be unreliable if cancer stages are used as measures of screening effectiveness or as surrogate markers for predicting an effect on breast cancer mortality.

What is the effect of screening?

Comprehensive systematic reviews have suggested that mammography screening reduces breast cancer mortality by 15-16% [1, 2]. This estimate is of the same size as the bias in the regression analysis of node-positive cancers.

Considering also the substantial bias related to determination of cause of death, the many flaws in the design and execution of the trials [1, 2] and the lack of an effect on all-cancer mortality, it seems reasonable to question whether screening has any life-extending effect [1, 2]. The present study and recent observational studies [39] support this concern.

CONCLUSION

The differences in the reported reductions in breast cancer mortality in the screening trials cannot be explained by differences in screening effectiveness. It is not clear what the effect of screening is, as the size of the bias was similar to the estimated effect.

CORRESPONDENCE: Peter C Gøtzsche, The Nordic Cochrane Centre, Rigshospitalet, Department 3343, 2100 Copenhagen Ø, Denmark.
E-mail: pcg@cochrane.dk

ACCEPTED: 6 January 2011

CONFLICTS OF INTEREST: none

FUNDING: not relevant

ACKNOWLEDGEMENT: I thank statistician *Per-Henrik Zahl* for his comments on the manuscript.

LITERATURE

- Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database of Systematic Reviews* 2009;4:CD001877.
- Humphrey LL, Helfand M, Chan BK et al. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002; 137:347-60.
- Vainio H, Bianchini F. *IARC Handbooks of Cancer Prevention. Vol 7: Breast Cancer Screening*. Lyon: IARC Press, 2002.
- Tabar L, Smith RA, Duffy SW. Update on effects of screening mammography. *Lancet* 2002;360:337.
- Moss SM, Cuckle H, Evans A et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet* 2006;368:2053-60.
- Miller AB, Baines CJ, To T et al. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *CMAJ* 1992;147:1459-76.
- Miller AB, Baines CJ, To T et al. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *CMAJ* 1992;147:1477-88.
- Miller AB, To T, Baines CJ et al. The Canadian National Breast Screening Study: 1. Breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med* 2002;137:305-12.
- Miller AB, To T, Baines CJ et al. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. *J Natl Cancer Inst* 2000;92:1490-9.
- Miller AB. The costs and benefits of breast cancer screening. *Am J Prev Med* 1993;9:175-80.
- Andersson I, Aspegren K, Janzon L et al. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *BMJ* 1988;297:943-8.
- Nystrom L, Rutqvist LE, Wall S et al. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993;341:973-8.
- Tabar L, Fagerberg CJG, Day NE. The results of periodic one-view mammographic screening in Sweden. Part 2: Evaluation of the results. In: Day NE, Miller AB (eds.). *Screening for breast cancer*. Toronto: Hans Huber, 1988:39-44.
- 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.
- Tabar L, Chen HH, Duffy SW et al. Primary and adjuvant therapy, prognostic factors and survival in 1053 breast cancers diagnosed in a trial of mammography screening. *Jpn J Clin Oncol* 1999;29:608-16.
- Frisell J, Lidbrink E, Hellstrom L et al. Follow-up after 11 years – update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res Treat* 1997;45:263-70.
- Frisell J. *Mammographic screening for breast cancer*. Stockholm: Södersjukhuset, 1989.
- Nyström L, Larsson L-G. Breast cancer screening with mammography. *Lancet* 1993;341:1531-2.
- Bjurstam N, Bjørneld L, Warwick J et al. The Gothenburg Breast Screening Trial. *Cancer* 2003;97:2387-96.
- Bjurstam N, Bjørneld L, Duffy SW et al. The Gothenburg Breast Cancer Screening Trial: preliminary results on breast cancer mortality for women aged 39-49. *J Natl Cancer Inst Monogr* 1997;(22):53-5.
- Bjurstam N, Bjørneld L, Duffy SW et al. Author Reply. *Cancer* 1998;83:188-90.
- Shapiro S, Venet W, Strax P et al. Ten- to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst* 1982;69:349-55.
- Shapiro S. Periodic screening for breast cancer. The HIP Randomized Controlled Trial. *Health Insurance Plan. J Natl Cancer Inst Monogr* 1997;(22):27-30.
- Roberts MM, Alexander FE, Anderson TJ et al. Edinburgh trial of screening for breast cancer: mortality at seven years. *Lancet* 1990;335:241-6.
- Alexander FE, Anderson TJ, Brown HK et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* 1999;353:1903-8.
- Moss S, Thomas I, Evans A et al. Randomised controlled trial of mammographic screening in women from age 40: results of screening in the first 10 years. *Br J Cancer* 2005;92:949-54.
- Shen Y, Yang Y, Inoue LY et al. Role of detection method in predicting breast cancer survival: analysis of randomized screening trials. *J Natl Cancer Inst* 2005;97:1195-1203.
- Tabar L, Fagerberg CJG, 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;1:829-32.
- Fagerberg G, Baldetorp L, Grøntoft O et al. 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.
- Tabar L, Fagerberg G, Day NE et al. Breast cancer treatment and natural history: new insights from results of screening. *Lancet* 1992;339:412-4.
- 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.
- Thomas LB, Ackerman LV, McDivitt RW et al. Report of NCI ad hoc pathology working group to review the gross and microscopic findings of breast cancer cases in the HIP study. *J Natl Cancer Inst* 1977;59:496-541.
- Moss S, Waller M, Anderson TJ et al. Randomised controlled trial of mammographic screening in women from age 40: predicted mortality based on surrogate outcome measures. *Br J Cancer* 2005;92:955-60.
- Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;339:b2587.
- Rapport över mammografiscreening i Kopparbergs och Östergötlands läns landsting (WE-projektet) – resultat efter första screeningsomgången. Stockholm: Socialstyrelsen, 1982.
- Miller AB. The Canadian National Breast Screening Study: update on breast cancer mortality. NIH Consensus Development Conference on Breast cancer screening for women ages 40-49. Kensington: National Institutes of Health, 1997:51-3.

37. Miller AB, To T, Baines CJ et al. The Canadian National Breast Screening Study: update on breast cancer mortality. *J Natl Cancer Inst Monogr* 1997;(22):37-41.
38. Gøtzsche PC. On the benefits and harms of screening for breast cancer. *Int J Epidemiol* 2004;33:56-64.
39. Jørgensen KJ, Zahl P-H, Gøtzsche PC. Breast cancer mortality in organised mammography screening in Denmark: comparative study. *BMJ* 2010;340:c1241.