ARTICLES

Quantitative Interpretation of Age-Specific Mortality Reductions From the Swedish Breast Cancer-Screening Trials

*Harry J. de Koning, Rob Boer, Peter G. Warmerdam, Petra M. M. Beemsterboer, Paul J. van der Maas**

Background: **Results from five Swedish randomized trials may provide the most conclusive evidence on the effect of mammographic screening and have been used to forecast the expected reduction in breast cancer mortality in other programs. However, those trials demonstrated different degrees of reduction. The interpretation of observed mortality reduction after long follow-up for women aged 40-49 years at trial entry is both important and controversial.** ***Purpose:*** **We estimated what percentage of the observed mortality reduction for women aged 40-49 years at entry into the five Swedish screening trials might be attributable to screening these women at 50 years of age or older. Moreover, we calculated the most likely percentage mortality reduction for specific screening programs if the Swedish results were generalized and analyzed whether characteristics of each trial might at least partly explain the observed differences in reductions among the trials.** ***Methods:*** **Each Swedish trial was simulated with one underlying computer simulation model (MISCAN—MIcrosimulation SCreening ANalysis) of the natural history of the disease and the performance of screening, taking into account nine important trial characteristics. Improvement in prognosis for screen-detected case patients was estimated with age-specific reduction for all trials and each trial design as a reference.** ***Results:*** **An expected 7% reduction in breast cancer mortality for women aged 40-49 years at trial entry (relative risk [RR] = 0.93) was determined by computer modeling, assuming no improvement in prognosis for cancers that are screen detected before 50 years of age. This result indicates that, of the overall 10% observed reduction (RR = 0.90) in the five Swedish trials analyzed, most (70%) of this reduction might be attributable to screening these women in later rounds after their 50th birthday. Using additional trial information, predictions of breast cancer mortality reduction in women 50 years or older might be 11% larger than previously expected, assuming that high-quality mammographic screening can be achieved in nationwide programs. For women aged 50-69 years at trial entry, the differences in expected versus observed mortality reduction among the trials are estimated to be relatively small. (Expected mortality reductions range from 24% to 32%.)** ***Conclusions:*** **Results from the Swedish randomized breast cancer-screening trials should be seen as more favorable regarding the effect of mammographic screening in reducing breast cancer mortality for women aged 50-69 years than was estimated earlier. Our analyses also suggest that the improvement in prognosis due to screening for women aged 40-49 years is much smaller than that for women aged 50 years or older. Approximately, 70% of the 10% observed reduction in breast cancer mortality (i.e., 7%) for women aged 40-49 years at trial entry might be attributable to a reduction due to screening these women after they reach age 50.** ***Implications:*** **Detailed screening data for the 40- to 49-year age group of all Swedish trials should be analyzed to specifically estimate the natural history and performance of screening in this age group. [J Natl Cancer Inst 87:1217-1223, 1995]**

It is still uncertain whether breast cancer screening for women under 50 years of age is effective in reducing breast cancer mortality (*1,2*). After eight rounds of biennial screening in a program generally considered to have reached good quality, the Nijmegen screening project recently showed no positive effect (*3*). According to some investigators, the results from the five Swedish randomized breast cancer-screening trials (Malmö, Kopparberg, Östergötland, Stockholm, and Göteborg) (*4-6*) can be considered to give the most conclusive evidence on the effect of mammographic screening. The published 10%-13% breast cancer mortality reduction rates for Swedish women under 50

**Affiliation of authors:* Department of Public Health, Erasmus Universiteit Rotterdam, The Netherlands.

Correspondence to: Harry J. de Koning, Ph.D., Department of Public Health, Erasmus Universiteit Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

See "Notes" section following "References."

years of age entered in a randomized study may seem encouraging. However, some women in this age group were also screened when they were 50 years old or older. Part of the observed mortality reduction in these women is likely to have been a result of detecting the cancer earlier in later rounds when the women were 50 years old or older, as was seen in the Health Insurance Plan (HIP) trial (*7*).

Furthermore, any trial is specific in its design, quality, and background situation. Consequently, different screening trials will result in different breast cancer mortality reductions in the study group as compared with the control group, even for the same age categories. The new and more detailed results from all five Swedish randomized trials, specified per age category (*4,5*), again show that the estimates of reduction vary widely between the trials. It is important to analyze to what extent differences in the characteristics of the five trials are likely to have caused these variations. The earlier publication of the seemingly less favorable results from the randomized trial conducted in Malmö led to a discussion of the efficacy of mammographic screening (*8*). The 21% reduction rate achieved for women in the invited age group inclusive of those aged 55-69 years seemed to be in contrast to the earlier published 39% rate in the group aged 50-74 years from the randomized trial in Kopparberg and Östergötland (*9*). Important characteristics of the trials, such as screening interval, attendance rate, follow-up period, and age groups, should be considered, however (*10*).

These characteristics of the trials are especially important with regard to screening in younger women. No individual trial has had the power to show a statistically significant mortality reduction in younger women. Much effort is being put into new trials for women under age 50, e.g., the U.K. trial, for which these issues are highly relevant. By using one underlying model that incorporates both the natural history of breast cancer and the performance of mammographic screening, we have analyzed all five trials and have taken into account nine important characteristics within each trial. The different policies for women aged 50 and above or under age 50 at entry are distinguished, as are characteristics in screening practice regarding intervention in the control groups. The goal is to adjust for as many relevant characteristics in screening policy as possible in each trial that may have influenced the outcome and estimate the improvement in prognosis for screen-detected cases. Although each trial may then be unique, the five Swedish trials should help in quantifying the breast cancer mortality reduction expected in other screening programs. Our analysis addressed three questions: 1) Which percentage of the observed mortality reduction for women aged 40-49 years at entry into the trial was likely to have been due to screening these women when they were 50 years or older? 2) What is the extent of breast cancer mortality reduction to be expected for present and future screening programs, if the Swedish results are generalized to, for example, those from The Netherlands and the United Kingdom? 3) Does the information (more details available) from the five trials at least partly explain the differences in observed mortality reduction and predict differences in efficacy between the five trials?

Methods

Underlying Model of Natural History of Breast Cancer and Performance of Screening

The computer simulation package MISCAN (MIcrosimulation SCreening ANalysis), developed at our institute, was used to evaluate the five screening trials, in which the natural history of the disease, the epidemiology, the design of the screening program, and the performance of screening are incorporated (*11,12*). The natural history of breast cancer is modeled as a progression through a number of stages. The first stage is no breast cancer; women are included in this stage until a transition occurs to one of the preclinical stages when a tumor becomes detectable by screening. There are one ductal carcinoma in situ (DCIS) stage and four invasive stages in the model, according to T-status (T1a, T1b, T1c, and T2+). The duration in the different stages follows an exponential distribution. The transition to the clinically diagnosed stages (with the same subdivision) is governed by the data on incidence and clinical stage distribution. In the case of early detection, women will enter the screen-detected stages (again with the same subdivision). The two end stages of the model are death from breast cancer and death from other causes, based on mortality data.

Key parameters in the model of the performance of screening are mean duration of screen-detectable preclinical disease, sensitivity, and improvement in prognosis for screen-detected cancers. Age-specific assumptions with regard to the mean duration of the five preceding screen-detectable preclinical stages of breast cancer and the sensitivity of screening had been validated with the use of all data from the Dutch screening projects (including women aged 35 years or over in the Nijmegen project and women aged 40 years or over in the Utrecht project) covering different periods and screening intervals (*10,13*). Sensitivity is stage dependent: 40%, 65%, 80%, 90% and 95% (aged 50 years or older) for DCIS, T1a, T1b, T1c, and T2+, respectively. The sensitivity for screening women under age 50 was estimated in the model to be 60% (ages 40-44 years) or 80% (ages 45-49 years) of the sensitivity for screening women older than 50 years. The mean duration of the preclinical screen-detectable period was approximately 1.8 years at age 35 and 6.2 years at age 70. These assumptions resulted in a good fit between model predictions and observed detection rates and interval cancers (both by age, stage, screening round, and interval) in the Dutch screening projects. (Detailed analyses of fit can be obtained from the investigators.) Other epidemiologic parameters (e.g., incidence, stage distribution, and mortality) were based on the Dutch data.

If one applies screening to a population, the shift from diagnosing relatively large clinical cancers toward diagnosing earlier (screen-detected) stages of cancer results in a decrease in breast cancer mortality, as shown in the randomized screening trials. In the model, women with screen-detected cancers can have a reduced risk of dying of breast cancer depending on the cancer size at detection. The degree of this improvement in prognosis after early detection had first been estimated on the basis of survival differences between women with screen-detected cancer and women with cancer diagnosed in the control group of the HIP trial (category-specific estimates based on differences in lymph node metastases), corrected for lead time (*12,14*). An additional amount of improvement had to be (plausibly) assumed to achieve the reduction by the model as observed in more modern trials, such as those conducted in Malmö, Kopparberg, and Östergötland (*10*). The number 1 minus the ratio of the risk of dying of screen-detected breast cancer divided by the risk when the cancer had been diagnosed in the absence of screening was estimated to be 0.80, 0.73, 0.51, and 0.35, respectively, for cancers screen detected in stages T1a, T1b, T1c, and T2+. Before the Swedish overview was published, we simulated the Malmö, Kopparberg, and Östergötland trials, using these assumptions and taking into account the differences in design. The simulated reduction in the breast cancer mortality rate (given the natural history, performance of screening, and the epidemiology as described before) was the same as the weighted average reported from these trials for women aged 50-69 years, under the assumption that sensitivity and mean duration of preclinical, screen-detectable disease did not differ strongly from the Dutch situation (*10,15,16*). Appendices on all the variables taken into account in the model are published in (*11,14*).

Swedish Trial Characteristics and Overview of Mortality Reduction

For the present analysis, each of the five Swedish screening programs was characterized, including intervention for the control group after a specific time

(as in the follow-up model). A review of all publications from the five trials, updated with personal information from the trialists, provided nine specific characteristics of all five trials: 1) age distribution at entry, 2) attendance rates for first round in the study group by age, 3) screening interval for study group by age and period, 4) start of intervention in control group, 5) attendance rates for first round in the control group by age, 6) assumed screening interval by age for women from the study and control groups after discontinuation of the trial (start intervention control group), 7) mean duration of follow-up, 8) woman-years per age and study and control category up to December 31, 1989 (Table 1). With these characteristics, for all women who entered the trials in either arm or either age group, both the number and timing of screens are determined on the basis of the mean follow-up period, the screening interval, and the attendance pattern in each trial. For example, the women in the study group of the Stockholm trial who had an average follow-up of 8.3 years and an average screening interval of 2.3 years will have had a maximum of four screens. Women in the control group were invited to participate in screening an average of 4.0 years after the start of the trial for the first time and may have had two screens in this analysis (8.3 – 4.0 divided by the 2.3-year interval). The attendance rates in some subsequent rounds are known, but the attendance rates in subsequent rounds of first attenders versus first nonattenders are published for only the Kopparberg, Östergötland, and Stockholm trials and were assumed to be the same in the other trials (characteristic 9). If no information was available, we assumed that, after the first screening of the control group, women had been invited to participate in screening up to December 31, 1989, according to actual age and according to the initial screening policies. The screening policy and, therefore, the number and timing of screens after age 50 for women who entered the studies at 40-49 years of age are certainly different in the intervention and control arms of the Kopparberg and Östergötland trials, since each trial had a different screening interval according to age at entry. Information was not available for all trials on the age-specific percentages of women in the control group who had mammography for nonsymptomatic reasons before being invited to participate in the screening. We have not assumed any difference in sensitivity with regard to one-view or two-view mammography.

The underlying model of the natural history of disease, the age-specific sensitivity and the age-specific improvement in prognosis were assumed to be the same for all the trials, and the expected case fatality (and breast cancer deaths differences between the study and control groups) was compared with the observed, for all trials together, taking into account the different designs of the five trials. The observed number of breast cancer deaths in each trial by age group at entry into the trial was based on the recent combined (independent) analysis of all causes of death among breast cancer patients in the trials, using the so-called follow-up model (*4*,*5*). By having detailed information on the number of deaths, screening policies, and designs for younger and older women, we could estimate the improvement in prognosis for screen-detected cases specifically per age category in the present analysis based on the recent overview. Since we had earlier estimates on improvement in prognosis based on the Malmö, Kopparberg, and Östergötland trials, we could compare these estimates with the new estimates. Finally, the expected consequences of the U.K. and Dutch screening policy were recalculated on the basis of the new information.

Statistical Analysis

Adaptation of the model was tested by using the sum of the squared differences between the observed and expected numbers of breast cancer deaths in each

Table 1. Characteristics of the five Swedish randomized breast cancer-screening trials used in predicting effect on breast cancer mortality reduction rates with one model of natural history of breast cancer and screening quality

Characteristic	Trial				
	Malmö	Kopparberg	Östergötland	Stockholm	Göteborg
Age at entry, y	45-69	40-74	40-74	40-64	40-59
Attendance of women invited for mammographic screening, first round, %					
<50	79	94	93	81	85
50-59	78	93	90	80	83
60-69	68	91	85	81	—
≥70	—	81	76	—	—
Screening interval, y (age range)	1.75*	2.0 (40-49) 2.8 (≥50)‡	2.5,2.0† (40-49) 2.5,2.8† (≥50)‡	2.3	1.5
Attenders' (attending previous round) attendance, %					
<50	85	94	93	95	87
50-59	85	95	91	95	88
60-69	85	91	85	95	—
≥70	—	80	72	—	—
Nonattenders' (not attending previous round) attendance, %					
<50	30	50	29	16	40
50-59	30	43	30	16	28
60-69	30	33	27	16	—
≥70	—	37	25	—	—
Intervention control group, y after randomization (age range)	—	6.8	7.0	4.0	7.0 (40-49) 5.0 (50-59)
Attendance rate, %, first-round control group					
<50	—	90	90	77	67
50-59	—	87	87	77	78
60-69	—	80	80	79	—
Assumed screening interval, y, control group/study group after stopping trials	1.75	2.0/2.8§	1.25/2.25§	2.3	1.5
Mean follow-up, y	11.8	10.2	9.7	8.3	6.2

*Average, depending on breast density.
†First round, subsequent rounds.
‡Only two screens ≥70.
§According to actual age.

of the five trials divided by the estimated variance. The expected number of deaths and variance were derived by assuming the total number of deaths as observed and a binomial distribution over the study and control groups. The outcomes are chi-squared distributions with four degrees of freedom. We used the Mantel–Haenszel method to calculate the average of the relative risks (RRs) modeled for the different trials. The model was fitted to the observed average values of 0.70 (ages 50-69 years) and 0.90 (ages 40-49 years). The precision of these values was equal to what was observed.

Results

Table 2 shows the observed number of breast cancer deaths and RRs in all study and control groups combined per age category as well as the expected numbers and RRs obtained from the model. In all variants, we used the same underlying model of the natural history of disease and performance of screening for each of the five trials (i.e., sensitivity and mean duration of preclinical, screen-detectable disease by age and stage). The simulation of the specific Swedish trial designs then leads to different numbers of screen-detected cancers in the different stages, with a consequent reduced risk of dying of breast cancer. Also, one identical parameter for improvement in prognosis was assumed for all trials. First, this stage- and age-specific parameter was estimated to make the results of the model for all trials together consistent with the newest observed overall 30% reduction rate for ages 50-69 years (***line a***). Table 2 also shows the results from the trials and the model for women aged 40-49 years (or 45-49 years) at the time of random assignment (***lines b-d***). The observed reduction rate in the group aged 40-49 years for all trials combined was 10%, a third of that in the age group 50 years or older. All screen-detected cases resulting from the design and the underlying model have been given the same improvement in prognosis (***line b***), equal to the one estimated for women above 50 years of age (as in ***line a***). With that assumption, we would have expected an overall 23% reduction in breast cancer mortality (RR = 0.77) for all women aged 40-49 years at trial entry, given the characteristics and the follow-up periods of all trials (***line b***). Assuming no improvement in prognosis for cancers that are screen detected before 50 years of age and the same improvement as in ***line a*** for cancers that are screen detected at age 50 years or more, a 7% mortality reduction between the invited groups and the control groups would have been expected at the end of follow-up of these five trials (***line c***). In other words, in a so-called pessimistic variant in which there is no benefit in the model for women whose cancer

Table 2. Woman-years and observed numbers of breast cancer deaths in study and control groups and RRs for all five Swedish trials combined, per age category, compared with expected ones with the model having different assumptions on improvement in prognosis for screen-detected cases for women aged 40-49 years

Woman-years* (× 1000)		Observed breast cancer deaths* (all 5 trials combined)			Expected breast cancer deaths (by computer model)		
Study group	Control group	Study group	Control group	RR	Study group	Control group	RR†
(a) Women aged 50-69 years entering trial; assuming an observed overall 30% reduction (RR = 0.70) in the risk of women aged 50-69 years dying of breast cancer, for all trials, and fitting this same reduction to the model (RR = 0.70) for consistency.							
911	725	281	312	0.70‡	275.3	317.7	0.70§
(b) Women aged 40-49 years entering trial; assuming an improvement in breast cancer prognosis equal to the one estimated for women aged 50 years or above as in a.							
428	350	84	75	0.90*,‖	79.5	79.5	0.77¶
(c) Women aged 40-49 years entering trial; assuming no improvement in prognosis for cancers that are screen detected before 50 years of age and improvement in prognosis equal to the one estimated for women aged 50 years or above as in a.							
428	350	84	75	0.90*,‖	86.7	72.3	0.93#
(d) Women aged 40-49 years entering trial; assuming improvement in prognosis for women aged 40-49 years with screen-detected cancer to be much lower than that estimated for women with screen-detected cancer at ages 50 years and above and the fitting of expected RR of the model to precisely that of the observed RR.							
428	350	84	75	0.90*,‖	85.3	73.7	0.90**

*Observed number of breast cancer deaths in each trial by age at entry based on recent combined, independent analyses of all causes of death in the trials using the follow-up model and trial end point of December 31, 1989 (*4,5*); women aged 70 years or older at trial entry were excluded and estimated from numbers of women at entry.

†Calculated by the Mantel–Haenszel method.

‡Average of published RRs for all five trials (0.72 for age group 50-59 years; 0.69 for age group 60-69 (0.72 + 0.69)/2 = 0.70).

§RR estimated to make the result of the model for all trials consistent with the most current observed overall 30% reduction (RR = 0.70) for ages 50-69 years. RR was estimated using the same stage and age-specific parameter for improvement in breast cancer prognosis for all five trials.

‖Observed reduction in the risk of dying of breast cancer in women aged 40-49 years for all trials combined was 10% (RR = 0.90), one third of that for the age group 50 years or older (RR = 0.70; 30% reduced risk of dying of breast cancer).

¶Expected overall 23% reduction rate in breast cancer mortality (RR = 0.77) for all women aged 40-49 at trial entry: 1) assuming the improvement in prognosis to be equal to the improvement estimated for women aged 50-69 years and 2) taking into account the characteristics and the follow-up periods of all trials.

#Expected 7% reduction rate in breast cancer mortality (RR = 0.93) for women aged 40-49 years at the start of the trials where there is no benefit in the model for women whose cancer was detected by screening before age 50 (compared with no screening). This 7% benefit must be a model-derived result for women whose cancer was detected at age 50 or over, where a reduction in risk of dying of breast cancer was assumed. This expected 7% reduction rate (RR = 0.93) indicates that most of the 10% observed reduction rate (RR = 0.90), seven of 10 (70%), might be attributed to screening these women in the later rounds when they were already 50 years or older.

**Expected overall RR adjusted to fit the observed mortality reduction for all trials combined (RR = 0.90). Improvement in prognosis for women aged 40-49 years with screen-detected cancer was assumed to be much lower than that estimated for women aged 50 years or older with screen-detected cancer.

was detected by screening before age 50 (compared with no screening), we expected in this analysis that the five trials would have shown a 7% reduction on December 31, 1989, for women aged 40-49 years at the start of the trials. This expected 7% benefit, therefore, must have been derived from the model for women whose cancer was detected at age 50 years or more, where we did assume a reduction in risk of dying of breast cancer. This result shows that most (seven of 10 [70%]) of the observed reduction might be attributed to screening these women in later rounds when they were already 50 years old or older. To simulate the observed mortality reduction for all trials combined (RR = 0.90), the improvement in prognosis for women with screen-detected cancer in this age group had to be assumed to be much lower than that estimated for women above 50 years of age with screen-detected cancer; in ***line d***, the parameter has been adjusted to fit the expected overall RR in this age category precisely to the observed RR.

Earlier data published in 1988 and 1989 had shown an RR of 1.15 (Malmö, ages 45-54), 0.79 (Malmö, ages 55-69), and 0.62 (Kopparberg and Östergötland, ages 50-69) between study and control groups (*8,17*). In earlier analyses when advising about the possible introduction of screening in The Netherlands (published in 1991), we had found no evidence for a difference in efficacy between these trials on the basis of screening policies. If adjustments were made for interval, attendance, follow-up period, and age groups, these characteristics were expected to compensate for each other (*10*). The then resulting 32% weighted (on size of confidence intervals) average reduction rate in breast cancer mortality for women aged 50-69 years who were invited to participate in the Malmö, Kopparberg, and Östergötland trials seems, however, better than the most recent results published in 1993 from all five Swedish trials, with 28% in the age group 50-59 years and 31% in the age group 60-69 years (average, 29.5%) used for this analysis (*4*). In the present analysis, however, the Göteborg and Stockholm trials are included with less broad age ranges, intermediate attendance rates, and other screening intervals. Furthermore, more detailed information with regard to the characteristics of the trials has become available. In fact, we now would have expected a smaller overall reduction for all five trials with the model of 26.5% (weighted on trial sizes) if the old estimate (*10*) on improvement in prognosis for screen-detected cases would still hold. The newest published results from all trials are, therefore, compatible with an 11% (29.5/26.5 times) larger improvement in prognosis for screen-detected case patients aged 50-69 years at randomization than previously expected.

The predicted breast cancer mortality reduction (not trial level) in the total population in The Netherlands with mammographic screening once every 2 years for women aged 51-69 years (Dutch nationwide policy) or once every 3 years for women aged 51.5-63.5 years (U.K. nationwide policy) can now be re-estimated, given the performance of screening in The Netherlands and this interpretation from the Swedish trials. The estimated improvement in prognosis on the basis of the five Swedish trials is used for the analysis of mortality reduction. As a consequence of our analysis, the predictions on reduction can be adjusted in a more favorable direction, based on the assumption that the high quality of screening can be achieved in such nationwide programs (*18*). In The Netherlands, a 17% reduction in the annual total female breast cancer mortality rate seems realistic (meaning >800 breast cancer deaths prevented per year). The expected reduction rate of 11% for the other practices (15% in the group invited to participate in the screening) is likely to be somewhat higher in the United Kingdom because of the presumably worse clinical stage distribution compared with that seen in The Netherlands.

For all trials together, we were able to make a good fit (agreement) between the observed reduction and the model. Table 3 shows the expected RRs for each trial and compares them with the observed RRs. In general, for women aged 50-69 years at trial entry, the differences in the expected mortality reduction rate between the trials are estimated to be relatively small (range, 24%-32% reduction), considering the different trial designs. It can be seen that the mortality reduction in this age group is expected to be the smallest in the Stockholm trial because of its specific characteristics. The Malmö and Göteborg

Table 3. Expected RRs in each trial per age category if improvement in prognosis is based on all trial results (top line), but individual trial characteristics as in Table 1 are taken into account, compared with observed RR (each trial assumed to have the same improvement in prognosis for screen-detected cases)

	50-69 y				40-49 y			
	Woman-years* (× 1000)				Woman-years* (× 1000)			
Trial	Study group	Control group	Expected RR	Observed RR* (deaths per study group)	Study group	Control group	Expected RR	Observed RR* (deaths per study group)
All	911	725	0.70†	0.70‡ (281)	428	350	0.90†	0.90* (84)
Malmö	193	193	0.68	0.86 (79)	46	47	0.82	0.51 (8)
Kopparberg	249	115	0.70	0.61 (79)	107	56	0.90	0.76 (26)
Östergötland	224	213	0.71	0.69 (69)	104	106	0.90	1.29 (24)
Stockholm	180	100	0.76	0.65 (33)	107	64	0.95	0.99 (20)
Göteborg	65	104	0.68	0.91 (21)	64	77	0.92	0.72 (6)

*Observed number of breast cancer deaths in each trial by age at entry based on recent combined, independent analyses of all causes of death in the trials using the follow-up model and trial end point of December 31, 1989 (*4,5*); women aged 70 or older at trial entry were excluded.

†Calculated by the Mantel–Haenszel method.

‡Average of published RRs for all five trials (0.72 for age group 50-59; 0.69 for age group 60-69; [0.72 + 0.69]/2 = 0.70).

trials' (with the shortest screening intervals) possibly could have been expected to lower the mortality relatively more than the other trials. In reality, the observed reductions varied more widely (9%-39%). Both the Kopparberg and the Stockholm trials have produced better RRs than expected. On the basis of our analysis, we expected a better RR for the Göteborg trial and the Malmö trial than was observed. The numbers, however, clearly were small, and the variance found in outcome was not larger than that expected from trials with the same design (chi-square = 5.05). Random fluctuation in the trial results is the most important explanation for the remaining discrepancies. Still, one can argue that differences in the nine important characteristics of the screening trials that were included in our analysis do not satisfactorily explain the observed differences in breast cancer mortality reduction between the five trials.

For women aged under 50 years at trial entry, Table 3 shows that the Malmö trial (which, for example, has a short screening interval, no official intervention in the control group, and the longest duration of the trial) would be expected to result in the largest breast cancer mortality reduction rate (RR = 0.82). The two most recently started trials, Göteborg and Stockholm, are expected to show only small reductions (RR = 0.92 and RR = 0.95, respectively). It is clear that the variation between the trials in the reductions (expected) is larger for this age group (range, 5%-18% reduction), again given the different designs and/or follow-up periods. The comparison between expected and observed results for this age group in each trial is, of course, strongly hampered by the small number of women invited to participate and the number of deaths expected (chi-square = 3.20). This situation would make a favorable conclusion about the Malmö trial for this age group still speculative. If there are no important details found from the Östergötland trial, other than those taken into account, that differ strongly from the other four trials, these are probably the best estimates on mortality reduction, including a strong negative effect from this specific trial. The Kopparberg trial has better RRs than expected for both young and older women.

Discussion

This analysis shows that the newest results from the Swedish randomized breast cancer-screening trials should be seen as more favorable with regard to the effect of beast cancer screening for women aged 50-69 years than earlier estimated (only from the reports from Malmö, Kopparberg, and Östergötland). Although the weighted average observed reduction reported presently is smaller than the earlier published average, our analysis shows that specific characteristics of all five trials are responsible for this. The longer follow-up, the additional details about the programs, and especially the information about dilution of the effect due to intervention in the control groups have been important. It is difficult to say whether the blind and uniform ascertainment of the deaths from breast cancer by an independent panel has influenced results, but for women aged 70 years or older and for women under age 50 years in the Östergötland trial, the absolute numbers now classified do differ strongly from the earlier published numbers (*4,5,17*). Our analysis also led to the assumption that the improvement in prognosis due to screening for women aged 40-49 years is much smaller than for women 50 years old or older. About 70% of the reduction observed in the trials' results for women aged 40-49 years at entry into the trial might be attributed to a reduction due to screening these women when they were 50 years or older. This fact should be borne in mind in the expectations for any trial on women under age 50 years. These results are important for a correct interpretation of the possible achievement of screening in present or future programs (*19*). Although any trial or program is unique, we think that a 25% reduction rate for the invited group aged 50-64 years in the U.K. screening program as estimated by other investigators (*20*) seems high, even if this overview analysis would have been based on the Kopparberg and Östergötland trials only. We believe that the present method provides a good (and maybe the last) opportunity for predictions to be made on mortality reduction at a nationwide level. The conclusion of the trials and the screening of the control groups will influence the differences in breast cancer death rates in both groups with longer follow-up. Also, to interpret the observed mortality differences in the future, analyses including the effect of intervention in the control groups will be required.

One should be cautious in interpreting the differences in this analysis between the observed and expected RRs in each trial. Without modeling, clearly, the numbers are so small that the variance in outcome found is not larger than expected from trials that have the same design. Random fluctuation in the trial results is the most important explanation for the discrepancies, but two other possible explanations are important. On the one hand, certainly a number of details from the trials or the baseline situation in the populations studied have not been available to us: regional epidemiology (population structure, incidence by age, stage distribution and treatment, and survival), details about the influence of the type of random assignment, and especially details about the situation after the first screening of the control group. Still, given the important details taken into account for each trial, this information is unlikely to strongly influence the present results and conclusions concerning women aged 50-69 years. On the other hand, other factors might have influenced the outcome of screening. Further research should be initiated to quantify the quality of screening, especially in younger women.

Such detailed information is not available in the literature on all five trials (*21*). Although it seems appropriate to base this analysis with respect to the natural history of breast cancer and age-specific sensitivity partly on Dutch data, where there are no data available to assume a reasonable difference between both countries (*13,15*), detailed information on detection rates, interval cancers, stage distributions, and the background situation in each Swedish trial are needed. It would then be possible to estimate whether our assumptions with regard to either natural history or sensitivity might have to be adjusted. The especially interesting question is whether it is indeed correct to attribute most of the published mortality reduction for women less than 50 years of age who were invited to screening above this age group. A standard meta-analysis does not account for the underlying parameters causing a possibly different RR in a particular screening situation.

A further analysis is needed with detailed Swedish screening data on women aged 40-49 years. Analysis of the other ran-

domized trials could be performed as shown in this article. Although such an analysis as ours could never replace actual data from the different randomized, controlled trials, it would be very meaningful. With longer follow-up, the Swedish trials will likely show a higher or statistically significant breast cancer mortality reduction for the women aged 40-49 years at trial entry, but the analysis shows it will be crucial to get a reasonable estimate with regard to the amount of reduction achieved for these women on the basis of screening in later rounds.

References

(1) Fletcher SW, Black W, Harris R, et al: Report of the International Workshop on Screening for Breast Cancer. J. Natl Cancer Inst 85:1644-1656, 1993

(2) Eckhardt S, Badellino F, Murphy GP: UICC meeting on breast-cancer screening in premenopausal women in developed countries. Int J Cancer 56:1-5, 1994

(3) Peer PG, Holland R, Hendriks JH, et al: Age-specific effectiveness of the Nijmegen population-based breast cancer-screening program: assessment of early indicators of screening effectiveness [*see* comment citation in Medline]. J Natl Cancer Inst 86:436-441, 1994

(4) Nyström L, Rutqvist LE, Wall S, et al: Breast cancer screening with mammography: overview of Swedish randomised trials [*see* comment citation in Medline] [published erratum appears in Lancet 342:1372, 1993]. Lancet 341:973-978, 1993

(5) Nyström L, Larsson LG: Breast cancer screening with mammography [letter]. Lancet 341:1531-1532, 1993

(6) Chamberlain J: Firmer evidence on the value of breast screening—the Swedish overview. Eur J Cancer 29A:1804-1805, 1993

(7) Shapiro S, Venet W, Strax P, et al: Ten- to fourteen-year effect of screening on breast cancer mortality. J Natl Cancer Inst 69:349-355, 1982

(8) Andersson I, Aspegren K, Janzon L, et al: Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. BMJ 297:943-948, 1988

(9) Tabár 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 1:829-832, 1985

(10) de Koning HJ, van Ineveld BM, van Oortmarssen GJ, et al: Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. Int J Cancer 49:531-537, 1991

(11) van Oortmarssen GJ, Habbema JD, van der Maas PJ, et al: A model for breast cancer screening. Cancer 66:1601-1612, 1990

(12) Habbema JD, van Oortmarssen GJ, van Putten DJ, et al: Age-specific reduction in breast cancer mortality by screening: an analysis of the results of the Health Insurance Plan of Greater New York study. J Natl Cancer Inst 77:317-320, 1986

(13) de Koning HJ: The effects and costs of breast cancer screening [dissertation]. Erasmus Universiteit Rotterdam, 1993

(14) van Oortmarssen GJ, Habbema JD, Lubbe JT, et al: A model-based analysis of the HIP project for breast cancer screening. Int J Cancer 46:207-213, 1990

(15) Day NE, Walter SD, Tabár L, et al: The sensitivity and lead time of breast cancer screening: a comparison of the results of different studies. *In* Screening for Breast Cancer (Day NE, Miller AB, eds). Toronto: Hans Huber Publishers, 1988

(16) Boer R, Warmerdam P, de Koning H, et al: Extra incidence caused by mammographic screening [letter]. Lancet 343:979, 1994

(17) Tabár L, Fagerberg CJ, Duffy SW, et al: The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. J Epidemiol Community Health 43:107-114, 1989

(18) de Koning HJ, Fracheboud J, Boer R, et al: Nation-wide breast cancer screening in The Netherlands; support for breast-cancer mortality reduction. National Evaluation Team for Breast Cancer Screening (NETB). Int J Cancer 60:777-780, 1995

(19) Beemsterboer PM, de Koning HJ, Warmerdam PG, et al: Prediction of the effects and costs of breast-cancer screening in Germany. Int J Cancer 58:623-628, 1994

(20) Vessey M: Breast cancer screening 1991: evidence and experience since the Forrest report. A report of the Department of Health Advisory Committee. NHS Breast Screening Programme. Sheffield, U.K.: NHSBSP Publications, 1991

(21) Tabár L, Fagerberg G, Chen HH, et al: Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. Cancer 75:2507-2517, 1995

Notes

Supported by the Health Insurance Executive Board in The Netherlands and the Bundes Minister für Forschung und Technologie, Federal Republic of Germany.

We thank the Health Insurance Plan and the Utrecht and Nijmegen projects for providing information for the earlier model on natural history and screening. We also thank L. Nyström and the Swedish trialists I. Andersson, N. Bjurstam, G. Fagerberg, J. Frisell, and L. Tabár for providing details about the designs and characteristics of each Swedish screening trial.

Manuscript received November 8, 1994; revised April 28, 1995; accepted June 9, 1995.