The benefit of population screening for breast cancer with mammography

Gezondheidsraad
Health Council of the Netherlands

Aanbiedingsbrief

The benefit of population screening for breast cancer with mammography

To:

the Minister of Health, Welfare and Sport

No. 2002/03E, The Hague, 7 March 2002

The Health Council of the Netherlands, established in 1902, is an independent scientific advisory body. Its remit is “to advise the government and Parliament on the current level of knowledge with respect to public health issues...” (Section 21, Health Act).

The Health Council receives most requests for advice from the Ministers of Health, Welfare & Sport, Housing, Spatial Planning & the Environment, Social Affairs & Employment, and Agriculture, Nature Preservation & Fisheries. The Council can publish advisory reports on its own initiative. It usually does this in order to ask attention for developments or trends that are thought to be relevant to government policy.

Most Health Council reports are prepared by multidisciplinary committees of Dutch or, sometimes, foreign experts, appointed in a personal capacity. The reports are available to the public.

The Health Council of the Netherlands is a member of INAHTA, the international network of health technology assessment (HTA) agencies that promotes and facilitates information exchange and collaboration among HTA agencies.

Preferred citation:
Health Council of the Netherlands: The benefit of population screening for breast cancer with mammography. The Hague: Health Council of the Netherlands, 2002; publication no. 2002/03E.

ISBN: 90-5549-420-8

Contents

Executive summary *9*

1 Introduction *13*

2 Breast-cancer screening in the Netherlands *15*

3 Methodological quality of the reviewed studies *19*
3.1 Criticisms made in the Cochrane review *19*
3.2 The committee's view *22*

4 Breast-cancer mortality as a an endpoint *31*
4.1 Criticisms made in the Cochrane review *31*
4.2 The committee's view *33*

5 Overall mortality and total cancer mortality as an endpoint *39*
5.1 Criticisms made in the Cochrane review *39*
5.2 The committee's view *40*

6 Overtreatment *43*
6.1 Criticisms made in the Cochrane review *43*
6.2 The committee's view *43*

7 Observational studies *49*
7.1 Case-control studies *49*
7.2 Effect evaluation *49*

8 Answers to questions posed by the minister *53*

References *55*

Annexes *63*
A Request for advice *65*
B The committee *67*
C Breast cancer mortality *69*
D All-cancer mortality *73*
E Overall mortality *75*

Executive Summary

In the autumn of 2001, the Cochrane Library and The Lancet published the results of a systematic review of randomised trials for early detection of breast cancer by mammography. The reviewers claimed that there was no reliable evidence to support the survival benefit of mammography screening. This has led to many discussions —both in scientific literature and in the lay press. The Dutch Minister of Health, Welfare and Sport requested from the President of the Health Council a rapid answer to the question of whether the outcome of the study, a so-called Cochrane review, nullifies the scientific basis of the current screening programme.

Service screening was gradually introduced, beginning in 1990. In 1999, the most recent year for which reports have been published, 744,000 women aged 50 – 75 years accepted the invitation to screening (78 per cent of those women invited). In 1996, when the screening programme had not yet covered the entire country, there were (according to the Dutch Cancer Registry) 4,400 women between the ages of 50 and 70 who were diagnosed with breast cancer. For half of these women the diagnosis was the result of screening. In 20 per cent of the cases this involved so-called interval cancer (breast cancer diagnosed in the time interval of two years between two successive screenings) and in almost 30 per cent of the cases it involved women who had never been screened. The President of the Health Council set up a committee that has compiled the present advisory report. As part of its work, the committee held a hearing that was attended by experts either involved in or opposed to screening, and re-examined the original studies.

The Cochrane review touches upon an important question: is breast-cancer mortality a valid endpoint for determining the efficacy of breast-cancer screening? This question is related to the ongoing debate on the design, analysis and methodological pitfalls of randomised trials of screening for (breast) cancer.

The Cochrane review is a systematic review (meta-analysis) of the results of published studies (randomised clinical trials or RCTs) into the benefit of population screening for breast cancer. The authors, two scientific staff members at the *Nordic Cochrane Centre* in Copenhagen, consider breast-cancer mortality to be an unreliable outcome. Of the seven eligible RCTs, they found that two were flawed; they were left out of consideration completely. Three of the RCTs were rated as having poor-quality data and the remaining two were rated as having medium-quality data.

The two RCTs of medium quality failed to find a statistically significant reduction in breast-cancer mortality in women who were offered screening. If the data from these two RCTs are combined with those from the RCTs for which the authors gave a quality rating of poor, then the results do, indeed, provide a statistically significant reduction. However, the authors consider breast-cancer mortality to be an unreliable outcome, biased in favour of screening.

On that basis, and assessed by overall (all-cause) mortality among the participants of the two medium-quality trials, they concluded that population screening has no survival benefit.

Furthermore, according to the version of the Cochrane review published by The Lancet, screening leads to increased use of aggressive treatment.

The committee endorses the quality criteria for the assessment of the eligible RCTs, but finds that they are inadequately specified and inconsistently applied. As a result, five of seven RCTs are left (completely or partially) outside the analysis and the trial results are weighted differently. Rather than keeping trials out of the analysis, the committee holds that it is better to use another method (sensitivity analysis) to investigate the effect of including or excluding data of lesser quality.

The committee agrees with the Danish scientists that the RCTs examined can be criticized in some respects, particularly in terms of randomisation. But, except for one trial, these shortcomings are not of a nature that renders unusable the published data. The committee does not find the reviewers' arguments convincing for scoring four of the RCTs much lower on methodological grounds than the two 'medium-quality' trials.

The committee considers as too extreme the conclusion that breast-cancer mortality is an unreliable outcome, biased in favour of screening. The Cochrane review

does indeed provide indications for possible sources of bias, but the authors do not provide evidence of important bias in favour of screening.

The committee does not agree with the conclusion that breast-cancer mortality as the primary endpoint must be replaced by overall mortality. They do, indeed, find that the use of breast-cancer mortality as the only outcome may cause one to overlook important harms (or benefits) of screening because of misclassification bias. Therefore, total cancer mortality, other important causes of death, and overall mortality must also be taken into consideration when interpreting the results of (breast-) cancer screening trials.

If screening has a beneficial effect on breast-cancer mortality, it should also be expected to have a (much smaller) beneficial effect on total cancer mortality and an (even smaller) effect on overall mortality. The reviewed RCTs were underpowered for detecting these small effects. Therefore, the requirement should not be that the small differences in question are statistically significant, but that they point in the right direction.

If data from all eligible trials (excluding Edinburgh) are taken into account for women older than 50 years, the relative risk for breast-cancer mortality is 0.72 (0.61 – 0.85) after 7 years, and 0.76 (0.67 – 0.85) after 13 years. In the same way, the relative risk for overall mortality is 0.97 (0.93 – 1.00) after 7 years, and 0.99 (0.97 – 1.02) after 13 years. The Cochrane review does not report the relative risk for total cancer mortality among women older than 50 years seperately from that among younger women.

That the RCTs showed no clear reduction in total cancer mortality (all women > 40 years) may be explained by the fact that breast-cancer mortality made up a small proportion of total cancer mortality (11 per cent in the Swedish RCTs). That is much lower than among women of the same age in the general population (24 percent in the Netherlands). This difference arises because women who were diagnosed with breast cancer before randomisation are, rightly, subsequently excluded from analysis because they cannot benefit from screening.

The committee is aware that screening causes an increase in the number of diagnostic procedures. Screening can also lead to treatment among women who would never have known about their breast cancer if it were not for screening, because they would have died from something else before the disease became clinically manifest.

The Cochrane review also draws attention to the possibility of a screening-associated increase in mortality. The authors predict that overall, radiotherapy is harmful for women at low risk of local recurrence, such as those identified by screening. As shown in RCTs carried out before 1975, radiotherapy after

mastectomy results in an excess of cardiovascular deaths. However, this risk is likely to be much lower with modern radiotherapy techniques. No vascular morbidity and mortality have been seen in the medium term (median observation period 10 years) with these techniques.

Screening detects smaller tumours which have not so often spread to the lymph glands. This change in stage distribution means that it is increasingly possible to use less mutilating surgery for these women. Furthermore, they will not as often need adjuvant therapy and regional radiation is less frequently used. The safety and effectiveness of radiotherapy and adjuvant therapy following breast surgery should be monitored (also in the long term) by research and periodic meta-analyses.

The committee finds it of crucial importance that well-balanced, honest advice be provided to the women involved regarding the risks and benefits of population screening. They urge the Ministry of Health, Welfare and Sport and the Health Care Insurance Board's National Coordinating committee for Population Studies to give the necessary attention to this.

The committee sees no scientific basis, in the light of the Cochrane review, to conclude that population screening for breast cancer for women over the age of 50 has no survival benefit. However, it does not rule out the possibility that new evaluations might show that the effect of screening on breast-cancer mortality is lower than was expected in 1990.

The committee therefore instructs that research should be conducted into the causes of declining breast-cancer mortality in the Netherlands. This study, which is already being prepared, will link at the individual level cause-of-death records with screening records and data on treatment. A solution must quickly be found to the problem that women who do not take part in screening cannot give consent to cancer registration, which would provide relevant data for this research.

The Cochrane Breast Cancer Group and the editor of The Lancet have argued for a full, independent systematic review based on individual patient data (IPD). This so-called IPD meta-analysis should also include updated outcome data, and should be revised on a regular basis in the light of new data. The committee supports this recommendation.

It is advised that a broadly diverse committee from the Health Council provides advice (in due time, when adequate new data are available) about the balance of risks and benefits of population screening for breast cancer. Updating of an advisory report presented by the Council in 1987 is in any case opportune, because improvements made since then in therapy, early diagnosis, and screening mammography all play roles in decreasing mortality from breast cancer.

Chapter 1

Introduction

Early in 2002, the Minister of Health, Welfare and Sport asked the President of the Health Council to urgently prepare a report on the scientific basis for breast cancer screening. The minister's request was prompted by the findings of a Cochrane review: a systematic review (meta-analysis) of randomised controlled trials (RCTs) of screening mammography run in various countries since 1963.

The Cochrane review, undertaken by two members of staff at the Nordic Cochrane Centre in Copenhagen, failed to find a decrease in overall mortality and breast-cancer mortality. Indeed, the reviewers claimed to have shown that population screening actually leads to increased use of aggressive treatment. These findings were published in two articles that appeared in The Lancet in January 2000 and October 2001 and in full on The Lancet's website (www.thelancet.com)(Göt00, Ols01, Ols01a). A modified version of the report was published by the Cochrane Library (Ols01b). In the latter publication – the official Cochrane review, although it does not represent an official Cochrane view of the usefulness of screening mammography (CBCG02) – the conclusion that screening led to more aggressive treatment was not included in the summary of the review's main results. Furthermore, by request of the editors of the Cochrane Breast Cancer Group it was added in the summary that if data of all eligible RCTs (excluding those conducted in New York and Edinburgh) are considered then the relative risk for breast-cancer mortality after 13 years is 0.80 (0.71-0.89) for women aged over 40. These text differences highlight the fact that there are differences of opinion as to how meta-analyses should be carried out and how the results should be interpreted. At another point, the summary stated that the evidence linking screening to

reduced breast-cancer mortality was inconclusive. This assertion was justified by the claim that breast-cancer mortality is an unreliable outcome.

The Cochrane Collaboration is an international non-profit organization that prepares, maintains and disseminates systematic reviews of the effects of health technologies. These activities are coordinated by fourteen Cochrane Centres, located around the world. Both the protocols for planned reviews and the finished reviews themselves are assessed by a Collaborative Review Group, in this case the Cochrane Breast Cancer Group.

On 28 January 2002, the President of the Health Council set up a committee to prepare a report in accordance with the minister's request. The text of the minister's letter is appended to this report (Annex A). The membership of the committee that produced this report is given in Annex B. The minister's key question was: Do the findings of the Cochrane review nullify the scientific basis for breast-cancer screening for women over the age of fifty?

The next chapter of this report provides a brief summary of the history and organization of breast-cancer screening in the Netherlands. In chapters 3 to 6, consideration is given to the main issues addressed by the Cochrane review, namely the methodological quality of the reviewed screening trials; the use of breast-cancer, total cancer and overall mortality to determine the effectiveness of screening; and the adverse effects of screening, such as overtreatment. In each case, first the Cochrane authors' comments are summarized, then the committee's assessment of these comments is presented. Particular attention is paid to the research involving women over the age of fifty, since the findings of such research underpin decision-making with regard to breast-cancer screening and its organization in the Netherlands. Chapter 7 is devoted to the results of observational (non-experimental) studies. The committee's response to the minister's enquiry is given in the final chapter.

Chapter 2

Breast-cancer screening in the Netherlands

In late 1974 and early 1975, observational studies on mammographic screening were started in Utrecht and Nijmegen. These studies were prompted by the encouraging early results of an American study involving women covered by the *Health Insurance Plan of Greater New York*. Started in 1963, the HIP trial, as it was known, was the first randomised controlled trial (RCT) designed to examine the efficacy of screening for (breast) cancer. The non-randomised projects in Utrecht and Nijmegen were evaluated as case-control studies (Col84, Col92, Ver84).

In line with recommendations made by the Health Council and the then National Council for Public Health (now the Council for Public Health & Care), the State Secretary for Public Health took the first steps towards establishing a national breast-cancer screeningprogramme in 1987 (GR87). The Health Insurance Funds Council (now the Health Care Insurance Board) was asked to finance and provide national coordination for the programme.

A national reference centre (the LRCB) was set up in Nijmegen, whose responsibilities included training radiographers, radiologists and pathologists, and monitoring the physical and technical quality of the mammographic screening. The professional associations agreed that targeted training and continuing education are a necessary precondition for qualitatively acceptable screening. They also supported the national guidelines concerning the diagnostic workup of mammographic abnormalities and concerning quality control of the screening. These guidelines were drawn up in 1988 under the auspices of the National Organization for Quality Assurance in Hospitals (now the Dutch Institute for Health Care Improvement) at the request of the

State Secretary. Pilot projects using either a mobile screening unit or a fixed screening unit were run in 1989.

Before a political decision was taken, a study was conducted to assess the likely benefits and risks, the consequences for the health care system, and the financial and staffing implications of a national screening programme. This study was carried out by an independent group of researchers who had not been involved in the Nijmegen or Utrecht screening projects. The conclusion was that, if women between the ages of fifty and seventy were screened for breast cancer every other year in the context of a national programme, the effects would on balance be positive, and that the benefits of such a programme would justify the cost (Kon90, Maa87). On the basis of RCTs conducted in Sweden (Kopparberg, Östergötland, Malmö), it was estimated that screening would reduce breast-cancer mortality among women between the ages of fifty and seventy by 33 per cent. The researchers expected that by about 2015 overall annual breast-cancer mortality could be reduced by 16 per cent. That would mean seven hundred fewer women dying of breast cancer each year than would otherwise have been the case (Kon90).

The national breast-cancer screening programme was introduced gradually, starting in 1990. In 1993, it was possible to invite a little over half the target group to screening; by the end of 1997, all women between the ages of fifty and seventy had been invited at least once (Kon00). From that point on, the programme was extended to include women between the ages of seventy and seventy-five.

Screening is organized regionally. There are nine screening regions, which coincide with the regions covered by the Comprehensive Cancer Centres. Each of these centres cooperates with the Municipal Health Services in its region to implement the programme. The LRCB in Nijmegen is responsible for medical and physical-technical quality control. Evaluation of the process and effects is undertaken by the National Breast Cancer Screening Evaluation Team (LETB), so that the programme can be modified or even cancelled as appropriate (Maa01). The LETB produces an annual report on the results of the screening programme (Fra00, Fra01). In addition, the team is currently making a detailed evaluation of breast-cancer mortality in the Netherlands to establish whether the anticipated reduction in mortality is actually beginning to occur.

Between 1990 and 1999, 5.7 million invitations were issued and 4.5 million examinations performed (a participation rate of 79 per cent). Suspicious abnormalities were observed by the radiologist assessing the mammograms in 45,600 cases (10.1 per thousand screening examinations). For 30,200 of the women involved (6.7 per thousand), further investigation involved a biopsy or another invasive procedure for the removal of tissue from the suspect part of the breast. In 21,500 of these cases (4.7 per thousand), breast cancer was diagnosed. In other words, more than seven in every

ten women ‘biopsied’ were found to have cancer. Some 24,100 women in whom abnormalities were observed were ultimately found to be clear of cancer; this equates to 5.4 false positive screening results per thousand mammographic screening examinations (Fra01).

The programme functions broadly as anticipated (Kon90, Maa87, Maa01). In some respects, things have actually worked out better than expected; the 79 per cent participation rate is well above the forecast 70 per cent, for example. There have also been less favourable results, however, such as a higher-than-expected number of women who were diagnosed with breast cancer within two years following a ‘negative’ screening result (i.e. in the interval between regular examinations): 0.99 per thousand woman-years, as opposed to the forecast 0.96 per thousand. One in three participants who contract cancer has an interval cancer. In 1996, half of all cases of breast cancer in the Netherlands involving women between the ages of fifty and seventy were screen-detected; just over 20 per cent had an interval cancer and nearly 30 per cent involved women who had not previously been screened. Concerns have been expressed regarding the communication between screening radiologists and diagnostic teams responsible for the evaluation of mammographic abnormalities (GR98, Hol00). If the benefits of screening are not to be lost, it is very important that this connection is well managed. At the minister’s request, the Health Care Insurance Board is looking into ways of improving matters in this area.

Generally speaking, there is a tendency for cancer screening to detect the more slowly progressive forms of a cancer (length bias sampling), some of which would not have become clinically significant (overdiagnosis bias), whereas interval cases will typically involve fast-growing, ‘aggressive’ cancer. Studies demonstrate that the survival rates from interval cancer are indeed lower than the rate among women whose breast cancer is detected by mammographic screening. The women involved in interval cases are no less likely to survive, however, than women who have never been screened (Bre95, Bur96, Col98, Hol86, Sch96, Sha82).

Disagreement regarding screening is nothing new. Criticism was voiced and down for discussion – in parliament and elsewhere – even when the Dutch breast-cancer screening programme was being set up (Kon90, Pee89, Sch90, WVC91). What is new is that an apparently sound meta-analysis should produce negative findings, since the results of earlier meta-analyses had been positive (Cox97, Cuc91, Fle93, Ker95, Nys93, Wal93).

Chapter 3

Methodological quality of the reviewed studies

A systematic review provides an up-to-date overview of previously published research findings in a particular field. Ideally, it should be based on the results of randomised controlled trials (RCTs). The pooling of statistically compatible data from different studies is known as a 'meta-analysis'. A meta-analysis may be based on published research results, but is more reliable if based on the individual patient data. A meta-analysis performed on the latter basis is referred to as an IPD meta-analysis. The Cochrane review now under consideration (Göt00, Ols01) was a meta-analysis. An example of an IPD meta-analysis would be a Swedish overview (Lar96, Nys93, Nys95, Nys96) of the RCTs performed in Sweden (An88, Bju97, Fri86, Tab85).

3.1 Criticisms made in the Cochrane review

The Cochrane review assessed seven eligible RCTs. These RCTs were started between 1963 and 1982 and all were completed more than ten years ago. The meta-analysis technique used by the Danish authors involved application of the following quality criteria:

- the randomisation is adequate and leads to comparable study groups
- post-randomisation exclusions are few or unbiased
- reliable outcome data are availble.

The authors assessed the RCTs on the basis of these criteria and placed them into four groups:

- high-quality trials: all criteria met
- medium-quality trials: only minor violations, important bias not suspected, or could be corrected
- poor-quality trials: major violations, important bias suspected and could not be corrected with available data
- flawed trials: major violations, important bias documented that could not be corrected.

Two RCTs (New York and Edinburgh) were considered flawed and therefore not suitable for use. This was because the intervention groups (those invited to screening) and control groups (not invited to screening) were not really comparable at the start of the trials or for analysis of the results.

On the basis of successive publications, the Danish authors calculate that in the HIP trial in New York 853 women were excluded from the study after allocation to the intervention group, while only 336 women were excluded from the similar-sized control group (31,092 women aged between forty and sixty-five). The women were excluded because of breast cancer diagnosed before their entry dates. The authors were also critical of the cause-of death (re)assessment. Both of these shortcomings were cited as evidence of bias in favour of screening.

The RCT in Edinburgh also involved women aged between forty and sixty-five, who were patients at 87 group practices. These practices were allocated randomly to the intervention and control groups, taking account of the number of GPs at each practice. This so-called ‘cluster randomisation’ led to a situation whereby the intervention group contained twice as many women of higher socioeconomic status (SES) as the control group (53 per cent, as opposed to 26 per cent).

The quality of three Swedish RCTs (Bju97, Fri86, Fri91, Tab85, Tab87, Tab89, Tab92) was rated poor by the Danish authors. The Two-County Trial (Kopparberg, Östergötland) was singled out for particular criticism because of an alleged lack of clarity regarding the cluster randomisation (which ‘may have been seriously flawed’), regarding the date of entry into the trial and regarding the number of participants (different numbers having been reported in the publications). The reviewers put forward various arguments to justify their assessment: breast-cancer mortality in the Kopparberg control group was higher than in the Östergötland control group (0.0021 compared with 0.0012, p= 0.02); there was an age difference at baseline between the intervention and control groups (+0.45 and -0.27 years, respectively) and the reports are unclear with regard to the exclusion of women. After calculating the discrepancy between the numbers of participants reported in two publications (Tab85, Tab89), the

Danish authors concluded that a disproportionately large number of women with previous breast cancer were excluded from the Kopparberg control group.

Two RCTs, one in Canada and one in Malmö (And88, Mil92, Mil00), were considered to be of medium quality. In both cases, potential participants were individually randomised. In the Canadian RCT, no women subsequently had to be excluded on the grounds of previously detected breast cancer, because this trial recruited volunteers, and all were asked about their history when invited to consent to participation in the trial. In Malmö, some women were subsequently excluded, but this was not felt to have affected the findings, since the exclusions were based on cancer registry data. Nevertheless, the Malmö RCT was not rated 'high-quality', because more women were excluded from the control group than from the intervention group and because the date of entry into the trial was defined differently for each group. No reason is given for not rating the Canadian RCT 'high-quality'.

The various RCTs produced quite different results, the reviewers note. Neither of the two 'medium-quality' RCTs revealed a significant decline in breast cancer mortality. By contrast, the 'poor-quality' RCTs found a marked effect. The difference between the 'medium-quality' and 'poor-quality' RCTs in terms of estimated effect is statistically significant. Heterogeneity of this kind is described by the Danish authors as very rare. They regard this as a strong warning signal that something is wrong, and claim a methodological explanation. The authors say that the results of empirical research into the methodological quality of RCTs indicate that RCTs exaggerate the estimated intervention effect by 33-41%, on average, if the randomisation procedure is not adequate or not described in the research report. The discrepancy between the effect estimated by the 'poor-quality' RCTs and that estimated by the 'medium-quality' RCTs is in good agreement with the latter finding.

The Danish authors find it disturbing that the strongest and most rapid decline in breast-cancer mortality were reported in RCTs that involved the least intensive screening (only two or three examinations offered, it is said, with long intervals), and where any contrast with the control group disappeared after only three to five years, they claim, because of the introduction of national screening. This is the opposite of what one would expect and suggests, say the reviewers, that differences in reported effects between the trials are attributable to a methodological artefact.

3.2 The committee's view

3.2.1 Assessment of the RCTs

The committee feels that the criticism of the HIP trial is partly justified. However, the alleged differential exclusion has been dealt with by one of the researchers involved (Mil01). After an interval (there having been no cancer registry), all deaths from breast cancer could be identified in the two groups. Identification of the date of diagnosis was then possible from hospital records, and patients diagnosed before randomisation were excluded. This procedure is unlikely to have biased the results.

In the Edinburgh trial, the randomisation procedure failed to stratify the group practices by socioeconomic status (SES). Consequently, the intervention and control groups were not comparable in terms of SES. Since there is a correlation between SES and both breast cancer risk and overall mortality (as well as between SES and screening attendance rate), the differences in SES are likely to have influenced the outcome of the RCT, as the researchers themselves noted (Ale89). A subsequent analysis was made, allowing for differences in SES (Ale99). From this analysis, it appeared that the original failure to allow for SES would tend to *under*estimate the effect of screening (Ale99).

The decision to also discount three Swedish RCTs is critical in relation to the review's conclusions regarding the benefit of breast-cancer screening. Where women over the age of fifty are concerned, the Cochrane review reports that the 'poor-quality' RCTs yield a statistically significant reduction in breast-cancer mortality by 31 per cent – RR = 0.69 (0.55-0.86) – but that the 'medium-quality' RCTs failed to show a significant reduction with a RR= 0.88 (0.64-1.20) after 7 years.

The authors of the Cochrane review conclude that the two categories of RCT are heterogeneous. This conclusion is based partly on a statistical test. Statistical heterogeneity, as it is known (i.e. significant differencences between effect estimates) may be attributable to medical heterogeneity (a difference in characteristics such as the nature and intensity of the intervention or the nature of the trial population, e.g. in terms of tumour size distribution), to methodological heterogeneity (a difference in methodological quality, relating, for example, to the comparability of the intervention and control groups obtained by randomisation, or to the temporal relationship between the intervention and the length of the follow-up period) or to some unknown coincidental phenomenon.

Inspection of the results reveals that the effect on breast cancer mortality in women aged fifty or older within an observation period of seven years estimated by pooling data from the two 'medium-quality' trials (RR = 0.88; 95 per cent confidence interval: 0.64-1.20) differs markedly from the effect estimated by pooling data from all six 'medium-quality' and 'poor-quality' trials (RR = 0.74; 95 per cent confidence interval: 0.62-0.89). A similar picture emerges if the estimate is based on an observation period of thirteen years. This is not the case, however, where the effect on total cancer mortality or on overall mortality is concerned. This fact might indicate that there were methodological shortcomings in the way breast-cancer mortality was measured in the 'poor-quality' trials. However, a cautious approach should be taken to interpreting these findings, since statistical heterogeneity may be attributable to any number of factors.

The various trials were, for example, medically heterogeneous in terms of the so-called 'control rate' – the stage distribution in the Canadian RCT control group being more favourable than that in the Swedish RCT control groups (Nar97). The average tumour diameter was 1.9 centimetres in the former and 2.8 centimetres in the latter, and the percentages of tumours measuring 2.0 centimetres or more were 30 and 60 per cent, respectively. No account was taken of these differences by the authors of the Cochrane review. Furthermore, medical heterogeneity could have arisen from a variety of other factors, such as differences in screening participation patterns, in the quality of the mammography, in the assessment of the mammograms, in the nature or quality of the diagnostic workup and therapy, in contamination of the control group or in effect divergence between women above and below the age of fifty. Any of these factors could also help to account for the observed differences between the RCTs in terms of the estimated effect of screening (Kon95). However, the fact that the trials differed where breast-cancer mortality was concerned, but not where total cancer mortality or overall mortality were concerned, would tend to suggest that medical heterogeneity is unlikely to be the sole cause of the statistical heterogeneity.

It is unlikely that methodological heterogeneity can fully account for the difference between the results of the 'medium-quality' RCTs and those of the 'poor-quality' RCTs. The methodological research that the Danish authors refer to (Kja01, Moh98, Sch95) does not adequately support their suggested explanation. The research in question does not provide evidence of a regular pattern, but highlights an association between shortcomings in the randomisation procedure and overestimation of the effect of intervention. Additionally, the committee shares doubts about extrapolation of the magnitude of the effects of inadequate randomisation from studies where control event rates are around 20 per cent to a screening trial where the event rates are far lower (Hay00).

Moving on to the question of randomisation in the reviewed RCTs , it should be said at the outset that randomisation at the individual level is generally preferable to randomisation at the group level, e.g. on the basis of geographical clusters. However, individual randomisation is not always possible or desirable. Cluster randomisation can sometimes be unavoidable for logistical or methodological reasons – if, for example, one is conducting a particularly large RCT or one needs to prevent contamination.

Three Swedish RCTs were cluster-randomised trials. With cluster randomisation, it is more likely that the intervention group and control group will differ in terms of prognostic characteristics such as age, because the number of randomisation units is smaller than would be the case with individual randomisation. In addition, the largest age imbalance – about five months in the Two-County Study – should be considered in the light of the fact that the trial involved women between the ages of 40 and 75 years. Therefore, such imbalances do not mean poor randomisation. They are not problematical, provided that the analysis is adequate (Cat00, Nix00, Sen01). If the presence of an imbalance can be defined, it is possible to deal with it by adjustment in the statistical analysis, just as in individually randomised trials. Where the 'poor-quality' RCTs are concerned, studies showed that such adjustment resulted in only marginal differences in the estimated effect of an invitation to screening (Nys02, Tab89).

Nevertheless, the reviewers used age as a marker for irregularities in the randomisation. Where the Two-County Trial is concerned, bias in favour of screening due to biased allocation is unlikely, since the age imbalance would tend to underestimate the benefit of screening. The comparability of the study groups has been examined in terms of all-cause mortality excluding breast cancer and in terms of all-cancer mortality excluding breast cancer. Those two basic outcome measures proved to be very similar (Tab88). The trial was designed partly with a view to obtaining socioeconomically comparable study groups. Statistical reanalysis of the Two-County Trial data by independent researchers has shown that this object was achieved (Nix00). Taking the cluster randomisation into account in the statistical analysis has little influence on effect estimation (Duf01, Nix00, Nys02, Tab92). In the overview of the Swedish RCTs, the cause-of-death pattern in the intervention group is, except for breast cancer, very similar to that in the control group, showing that the groups were comparable (Nys96, Tab89).

The higher level of breast-cancer mortality in the Kopparberg control group than in the Östergötland control group as indicated in the Cochrane review, is indeed notable. It does not automatically follow, however, that " the randomisation procedure may have been seriously flawed". Slightly different , geographic-region based, cluster randomisation methods were used in the two counties (Tab85, Tab92). It may be that

the difference in mortality reflects true regional variations in breast- cancer mortality. In this context, it should be noted that breast-cancer mortality in the Malmö control group was also higher than that in the Östergötland control group (Nys93). This explanation is in line with the observation that the stage distribution among women with breast cancer in the Östergötland control group was more favourable than among their counterparts in the Kopparberg group (Tab89). The committee does nevertheless concede that the Swedish trialists should have been able to provide a more convincing explanation, perhaps by reference to regional incidence and mortality data.

The RCT reports published by the Swedish trialists do partly explain the discrepancies in the numbers of participants in the course of the trial. The main factors cited are the exclusion from the trial of women who had been found to have breast cancer before the date of randomisation and differences in the way age is defined (on the basis of year of birth, or on the basis of exact date of birth) (Duf01, Nys93, Nys02, Tab89). A woman already known to have breast cancer cannot benefit from screening and should not therefore be considered when assessing the value of screening. In a trial based on pre-randomisation (where members of the control group are not asked to consent to inclusion in the trial), it was inevitable that some of the women involved would already have breast cancer at entry of the trial. When it became possible to 'clean up' the trial population by record linkage to cancer registry data, this was done and duly reported (Tab88, Tab89); subsequently published findings related only to the 'clean' study groups (Duf01, Tab89).

If one calculates the number of exclusions from the Two-County Trial in the way the Danish authors indicate, it does indeed appear that the exclusion rate for the Kopparberg control group (264 / 18,846 = 1.4 per cent) is a little higher than that for the intervention group (462 / 39,051 = 1.2 per cent). The committee does not believe that this is attributable to any difference in the thoroughness of the steps taken to identify women with breast cancer diagnosed before the randomisation date, since record linkage to cancer registry data was used in both cases. In the committee's view, the assertion that the discrepancy in the exclusion rate is attributable to irregularities in the randomisation procedure is unsubstantiated, but cannot possibly be ruled out completely. It is worth noting, however, that the discrepancy tends to result in *under*estimation of the benefit of screening.

3.2.2 *Application of the quality criteria*

It is not immediately apparent to the committee why three of the Swedish RCTs should be regarded by the Danish authors as 'poor quality', while two other trials (Malmö and Canada) are 'medium quality'. The committee therefore doubts whether the quality criteria have been consistently applied. In early 2000, the Danish authors suggested

that age imbalances between the intervention and control groups provided the main evidence of poor randomisation (Göt00). They calculated that the age imbalances in the 'poor-quality' RCTs varied from - 0.18 years (Stockholm) to +0.45 years (Kopparberg). Their initial conclusion that the age distribution was 'extremely skewed' and thus incompatible with reliable randomisation is not repeated in the Cochrane review, since it was rightly criticized by various commentators, including the Cochrane Breast Cancer Group (CBG02). The authors nevertheless stood by their assessment of the RCTs in question as 'poor quality', even though the trials' exclusion rates were neither particularly high nor biased (criterion 2), and even though either assessment or reassessment of causes of death was blind in each case (criterion 3).

Other considerations raise further doubt regarding the consistency of the application of the quality criteria. First, the Danish authors did not have data regarding the age distribution of participants in the RCTs in Malmö and Canada, yet this did not prevent them concluding that the randomisation procedures in these RCTs were unbiased (Göt00). The committee regards this as odd in view of the fact that other RCTs were described as 'very likely flawed' on the basis of quite modest age imbalances (Ols01).

Furthermore, despite the fact that the Canadian RCT, like the Malmö RCT, was less than perfect, the Danish authors attach less importance to the majority of its shortcomings. The issue of blindness was critical in the Canadian NBSS-1 because randomisation came after clinical examination of the breasts. Although an independent review (Bai97) has confirmed the integrity of (certain aspects of) the randomisation procedure used in the CNBSS-1, various commentators have raised doubts about this procedure, because of the study findings that showed an imbalance in the number of women with advanced breast cancer. In particular, the mammography group had an excess of women with involved lymph nodes whose breast cancer had been detected by physical examination at baseline (i.e., before randomisation), whereas the control women have rejoyed a remarkable survival, their death rate being, inexplicably, a little more than half the rate predicted (Boy97, CETS93, Duf00, Kop97, Met93, Tar95). The Danish team, however, accepted the integrity of the randomisation procedure without publishing their reasoning. In their explanation of the baseline imbalance in stage distribution they note that mammography group patients were generally treated in centres with more thorough axillary-node dissection, but did not deem this difference a source of bias. However, they wrongly cite this treatment difference as a potential source of bias in the other trials (Duf01, Hol86).

Criticism can additionally be levelled at the Cochrane review for its failure to make proper allowance for differences between the design of the Canadian RCT (with a volunteer study population) and that of the Swedish (population based) RCTs, in which participants were either offered mammographic screening or not offered

screening at all. Furthermore, all participants in Canada (CNBSS-2) were given a physical breast examination once a year and instructions on self-examination. Although it has not been shown that regular physical examination or self-examination is effective in reducing breast cancer mortality, it is not unreasonable to suppose that the additional emphasis on the early detection of breast cancer might have reduced the contrast between the two study groups. Stronger contrast reduction is likely to have resulted from the relatively favourable starting situation (tumour stage distribution, see above) with regard to breast cancer in Canada (Nar97), which will have made it harder to show the efficacy of screening. Because of the more favourable stage distribution, it was only to be expected that any effect on breast-cancer mortality would be less in Canada, and would take longer to manifest itself than in Sweden.

The reviewers also appear to have attached more importance to post-randomisation exclusions in some trials than in others. According to the review, the rate of exclusion from the Malmö RCT's intervention group due to pretrial breast cancers was 2.6 per cent (547 / 21,242), whereas the rate of exclusion from the control group was 2.2 per cent (457 / 21,240). Thus, exclusion was twice as common in the Malmö trial as in the Kopparberg part of the Two-County trial. Furthermore, differential exclusion in Malmö was more likely to influence effect estimation, and any such bias will have been in favour of screening. This is accepted as unbiased, because exclusion in the Malmö trial involved an official cancer register, an independent source. Yet the Cochrane review criticizes the way subjects were excluded from the Two-County Trial, despite the fact that the same procedure and data from the same (Swedish) cancer registry were used (Duf01, Nys93, Tab88, Tab92).

The Cochrane review suggests that the various RCTs' estimates of the effect of screening on breast-cancer mortality are inversely proportional to the intensity of the screening – the opposite of what one would expect. At first sight, this appears to be an important observation. However, it fails to take account of the optimal temporal relationship between intervention and follow-up duration for the anticipated effect, as established from accurate data (Mie02). Breast-cancer mortality in Malmö, for example, first began to decline after seven years of screening (And88), but no such decline was observed in the CNBSS-2, where screening continued for only three or four years (Mil00). Further research would be beneficial in this context. It should be noted that the Cochrane review misrepresents the screening intensity of the 'poor-quality' trials, underrating the number of screening rounds and the trial time.

3.2.3 Conclusion

When defining the quality criteria for their systematic review, the authors attached the greatest priority to the (internal) validity of the RCTs. In principle, the committee

supports this approach. However, the criteria definitions were not sufficiently explicit to minimize the danger of subjectivity in the quality assessment. The concept of 'adequate randomisation' is not properly defined, for example. Nor is 'reliable outcome data'. The way such concepts are interpreted has far-reaching consequences for the findings of the Cochrane review.

The committee is critical of the way the authors chose the quality criteria themselves, modified them during the course of the review and failed to apply them consistently. The consequence of this approach was that five out of seven RCTs were entirely or partially excluded from the analysis, when it was not necessary to do so. Given that all large-scale trials have strengths and weaknesses, the committee feels it would have been more appropriate to establish by means of sensitivity analysis whether each potential bias source was likely to lead to underestimation or overestimation of the effect of intervention and to determine the quantitative significance of the bias (CBCG02).

The committee believes that the decision to dismiss the HIP trial as flawed was unjustified, even though the trial does not carry great weight. On the other hand, the committee regards the baseline imbalance in SES in the Edinburgh RCT as too great for the findings to be taken into consideration, even after adjustment for SES of group practices.

It is worth noting that the HIP trial findings – a reduction in breast-cancer mortality of 22 to 35 per cent among women over the age of fifty (Ols01) – were not taken into account when the effect magnitude of introducing screening to the Netherlands was forecast in 1990 (Kon90). There were various reasons for this. First, mammography techniques had improved considerably since the HIP trial (when film was used without a screen). Second, the starting position with regard to breast cancer (tumour stage distribution and survival rates) in the USA during the 1960s was very different from that in the Netherlands in the 1990s. And, third, the screening in the HIP trial involved both mammography and physical examination.

Nor were the Edinburgh findings – a statistically non-significant reduction in breast-cancer mortality of 12 to 19 per cent among women over the age of fifty (Ols01) – taken into account in this context. The reasons being the low participation rate, the initially poor quality of the (xero)mammography and the combined use of physical examination and mammography screening (Kon90).

The committee rejects the contention that the inconsistency between the results of the 'medium-quality' RCTs and those of the 'poor-quality' RCTs is attributable solely to the methodological shortcomings of the latter. It is nevertheless accepted that such

shortcomings might have influenced the estimation of breast-cancer mortality effects to a limited extent.

Chapter 4

Breast-cancer mortality as a an endpoint

4.1 Criticisms made in the Cochrane review

Knowing whether a subject belonged to the intervention group (those invited to screening) or to the control group may effect the assessment of cause of death. This source of information or ascertainment bias can be reduced by blind assessment (i.e. assessment made without such knowledge). However, only the trials from Canada and Malmö (And88) made use of blind cause-of-death assessment (Göt00).

The Cochrane review concludes that breast-cancer mortality was an unreliable outcome, even when the cause of death was assessed blindly. In support of this verdict, the authors make a number of points. First, figures are presented (derived from Tab88) to illustrate directly that, in the Two-County Trial, death from non-breast cancer among women with a diagnosis of breast cancer was 2.4 times higher in the intervention group than in the control group. This increased mortality, the review claims, amounts to half the reported reduction in breast-cancer mortality.

Second, the authors draw attention to the issue of radiotherapy. Radiotherapy reduces the rates of local recurrence by two-thirds (EBC00). Early cancers are treated by tomourectomy and radiotherapy. The implication of this, according to the Cochrane review, is that deaths among women whose breast cancer was detected by screening are more likely to be misclassified as deaths from other causes and that too many deaths in the control group will be classified as breast cancer deaths.

Third, on the basis of data from the Malmö RCT (And88), it is pointed out that 21 per cent of women with breast cancer who died had or had previously had at least

one other malignancy, creating considerable potential for misclassification of the cause of death.

Finally, the Cochrane review suggests, belief in the effectiveness of a particular intervention may influence quite substantially cause-of-death assessment. In this context, the authors cite an American study (New00). This study showed that prostate cancer patients who did not die from prostate cancer had a proportionate mortality for other causes, which was very similar to a cohort of patients with benign prostatic hyperplasia. However, those with no initial treatment had a proportionate mortality for other cancer causes of 14%, significantly lower than the 19% seen in the non-prostate cancer cohort. Conversely, the cancer proportionate mortality in the "aggressively" treated subgroup was 30%, significantly higher than that in the non-prostate cancer cohort. This indicates, the Danish reviewers suggest, that if it is believed that an intervention is successful, there is a tendency to put some other cause of death on the death certificate.

Blind (re)assessment of the cause of death, the Cochrane reviewers argue, does not remove the objections to the use of breast-cancer mortalityas an outcome measure. First, because it is biased in favour of screening (due to differential misclassification). To support this contention, the authors refer to the results of the blind reassessment of cause of death in 144 dubious cases in the HIP trial. Differential misclassification, it is suggested, might be responsible for approximately half of the reported mortality benefit.The authors also claim to have found evidence of differential misclassification in the overview of the Swedish RCTs. Blind reassessment of the cause of death resulted in the conclusion that 418 women in the intervention groups (with a total observation period of 1.43 million woman-years) had died of breast cancer, while the corresponding figure in the control groups (1.14 million woman-years) was 425. On the basis of the officially recorded causes of death, the figures were 419 and 409, respectively. The fact that the net seventeen reclassifications was in favour of screening is regarded by the Danish authors as evidence of bias.

The authors additionally argue that, in the event of someone who has been screened subsequently dying (perhaps years later) as a result of the complications of some diagnostic or therapeutic intervention made in the context of screening, it is quite possible that no link will be made between screening and death. In this context, the finding that radiotherapy following mastectomy is associated with increased vascular mortality (ECBC00) is cited as evidence for bias in favour of screening, assuming higher rates of post-mastectomy radiotherapy among screened women.

4.2 The committee's view

Debate regarding the design, analysis, methodological pitfalls and reporting of RCTs to estimate the effect of screening or primary preventive programmes has been ongoing for some time (Bla02, Cox97, Juf02, Lar96, Mie02, Pac97, Sch90, Sha90). This is not particularly surprising, since population-based trials of this kind can never be perfect. What is more, the benefits of screening are often modest and therefore difficult to demonstrate. It is also possible for screening to do more harm than good. In view of these problems, RCTs need to involve very large numbers of subjects if they are to provide reliable findings. The Cochrane review illustrates just how difficult it is to carry out and analyse such trials.

The drawback of using cause-specific mortality, e.g. breast-cancer mortality, as an outcome measure is that it provides information only about the benefits of screening, not about any possible harm. In that sense, it is a surrogate outcome, which assumes that a reduction in cause-specific mortality must imply a net increase in the chance of survival (Juf02).

Furthermore, the use of cause-specific mortality as the primary endpoint renders screening trials subject to serious bias (Bla02). This criticism cannot be levelled at the use of overall mortality (i.e. all-cause mortality), which can provide unambiguous evidence of any increase in the chance of survival and help to highlight any harm from screening.

It is therefore pertinent to ask whether overall mortality should always be used as a primary endpoint. And whether screening RCTs should involve even more subjects than, for example, the 200,000 men that the International Prostate Screening Trials Evaluation Group hopes to include. Should the RCTs looking at breast-cancer screeninghave been five times as big, involving 2.4 million women, as stated in the Cochrane review? Or is such an approach statistically unnecessary, since the available numbers are large enough for getting stable estimates of the reduction or increase in overall mortality or total cancer mortality?

4.2.1 Breast cancer as a cause of death

The committee considers it appropriate to scrutinize the validity of certified causes of death. However, the accuracy of death certification is higher for malignant neoplasms than for other causes of death, and higher for breast cancer than for other cancers (Alb00, Ede99, Kir85, Pen01, Sat98). Research has shown that the chance of breast cancer being misclassified as the cause of death ranges from less than 5 per cent to a

maximum of 10 per cent (Ald67, And88, Bri84, Bro93, Cha91, Col92, Dijc96, Gar96, Nys95).

If breast cancer has metastasized and the woman dies, it is difficult to make a mistake when assigning the cause of death. The major difficulty is whether breast cancer was the cause of death when the deceased woman had been diagnosed with another malignant disease (Mil01). The Cochrane review suggests that such confusion is quite common, citing data from Malmö in support (And88). These data indicate that 21 per cent of breast cancer patients who die also have another (primary) form of cancer. This figure probably includes second (primary) breast cancers and non-melanoma skin cancers. Such malignancies should be excluded in this context, since co-morbidity of this kind cannot lead to uncertainty regarding the cause of death. Other studies yield a much lower percentage than 21, namely 6 to 12 per cent (Cha91, Nom99). Results of studies on the development of second (primary) tumours following treatment of breast cancer also point to a much lower percentage, in the region of 5 per cent (Fow01, Mat00, Obe00, Rub00). Taking these findings into consideration, the committee believes that the number of dubious cases is modest and does not create considerable potential for bias.

Not all commentators accept that belief in the effectiveness of an intervention can explain the findings of the American retrospective cohort study on causes of death in elderly men with prostate cancer (New00). The doctors involved were not asked about their preferred therapies. Besides, it is inevitable that one doctor will advocate ‘watchful waiting’, while another believes in radical surgery and a third in radiotherapy. Furthermore, the doctor who determines the cause of death is not always the one who previously decided what form of therapy the patient should receive – not least because ten years could elapse between diagnosis and death (New00). Another weakness of the research is that the certified causes of death were not assessed against a ‘golden standard’ (autopsy reports or all relevant clinical data). As the researchers stated themselves (New00), their work was designed to be hypothesis-generating. The research was not intended to test the validity of the information bias hypothesis; this hypothesis emerged in the course of post-hoc subgroup analysis. It is an important hypothesis, but one that has yet to be confirmed.

The findings of research into the validity of colorectal cancer death certification among participants in an RCT on the effectiveness of screening for colorectal cancer (Ede99) does not support the hypothesis of information bias.The discrepancy in gross number of deaths from colorectal cancer – 318 from the death certificates and 322 by the deaths review committee, a difference of 1,5% - was small and cannot have a major influence on the estimated effect of screening.

4.2.2 *Bias in the RCT findings*

Although the validity of breast cancer as the certified cause of death is high, it does not automatically follow that breast-cancer mortality as an outcome measure in experimental or observational studies is not subject to bias. The Cochrane review's conclusion that breast-cancer mortality is unreliable and biased in favour of screening, is based on the assumption that flaws are due to differential exclusion of women with previous breast cancer from analysis and differential misclassification of cause of death, which cannot be corrected for by blind (re)assessment of the cause of death.

As indicated in subsection 3.2.1, the suspicion of differential exclusion in the HIP trial was refuted by one of the researchers involved (Mil01). Also, exclusion in the Swedish RCTs was performed objectively, by record linkage to the Swedish Cancer Registry (see section 3.2). Furthermore, the committee cannot find any definite evidence in literature of *differential* misclassification of breast cancer as the cause of death (And88, Cha91, Dijc96, Gar96).

The reviewed RCTs do offer considerable scope for differential misclassification. The Cochrane review suggests that there is definite evidence of such misclassification occurring (see the first example given in section 4.1). This is a serious issue, since, if proven, the resulting bias would account for half of the reduction in breast-cancer mortality reported in the Two-County trial.

However, the committee believes that there has been a misunderstanding. The reviewers used data from the Two-County trial (Tab88) to show that women in the intervention group in whom breast cancer was diagnosed (1,295 as compared to 768 women in the control group) had a higher mortality from other malignancies (25 and 6, respectively) and also from all causes other than breast cancer (81 and 34, respectively) : RR = 2.4 and 1.4, respectively. The alleged excess rates in the intervention group can, however, at least partially be explained by the fact that screening will result in breast cancer being diagnosed several years earlier than would otherwise be the case. Because of this 'lead time', women with screen-detected breast cancer were at risk of death several years longer. The Cochrane review does not take account of this lead time bias. To make a valid comparison, the denominator should not be the number of women with breast cancer, but the number of woman-years at risk within the breast cancer cases. When comparison is made on this basis, no statistically significant discrepancy emerges between intervention and control groups in the Two-County Trial in terms of mortality from all causes other than breast cancer (i.e. overall mortality minus breast-cancer mortality) (Tab89, Tab92). The publications are, however, inconclusive with regard to mortality from other forms of cancer among women with breast cancer. More information on all-cancer mortality may be important,

since the Swedish overview indicates the presence of (statistically non-significant) excess mortality for gastrointestinal cancer in the intervention groups (Nys96).

The Cochrane review rightly states that blind (re)assessment of cause of death – no matter how careful – does not preclude the possibility of bias. Consider the Swedish overview, which was based on the original research data (concerning the individual participants in all Swedish RCTs), including the clinical data. These data were linked to cancer and cause-of-death registry data. A blind reassessment was made of the cause of death in all cases involving participants who died before 1990, having (according to the cancer registry) been diagnosed with breast cancer after the randomisation date. For the purposes of this blind reassessment, the relevant medical records were collected and blinded concerning identity and allocation status, and the cause of death checked by an oncologist. The reassessment was made by a committee consisting of a pathologist, a radiologist, a surgeon and a (second) oncologist, none of whom had been involved in the RCTs. Each committee member reassessed all 1,296 cases, without knowing the screening status of the deceased or their colleagues' opinions.

Other commentators have recently highlighted the possibility of so-called 'slippery-linkage bias' (Bla02). This form of bias tends to result in overestimation of the effect of screening (Bla02). Such bias occurs where a 'positive' screening result leads to an invasive or otherwise hazardous intervention, ultimately causing the subject's death, without this being associated with the screening. The Danish authors claim that bias of this kind is associated with a greater use of radiotherapy in screened women than in controls, citing the finding that post-mastectomy radiotherapy leads to excess vascular mortality (EBC00). Setting aside the issue of the safety of *modern* radiotherapy techniques (see also section 5.2), this underlines the importance of using other outcome measures in conjunction with disease-specific mortality when interpreting the results of cancer screening trials.

The contention that blind reassessment of the cause of death in dubious cases biased the results of the HIP trial is refuted by one of the researchers (Mil01). In his reply, Gøtzsche continued to maintain that there had been bias, pointing to the fact that there had been no decline in all-cancer mortality (Göt01). The committee takes issue with Gøtzsche on this point, however. According to the Cochrane review, the HIP trial yields a relative risk for all-cancer mortality of 0.98 (0.89-1.08) for women aged between forty and sixty-five (Ols01). In the committee's view, a reduction in all-cancer mortality of this order is consistent with a reduction in breast-cancer mortality of 17 per cent after thirteen years (see also section 5.2). As is the 1 per cent fall in the overall mortality: RR = 0.99 (0.94-1.05) after thirteen years. Therefore, the committee feels that serious bias is unlikely.

In the Swedish overview, there does appear to have been differential misclassification in favour of screening, if one looks only at the numbers and the direction in which classifications are shifted, and if one accepts the Swedish Cause of Death Register as the gold standard. Blind cause-of-death reassessment leads to some revision of the estimated effect: a reduction in breast-cancer mortality by 23 per cent, rather than 20 per cent (Nys95).

Several other studies with data from the Swedish overview have shown as well that only marginal differences between the relative risk estimates emerged when using different outcomes measures (Nys93, Lar96). One of the comparisons made use of 'breast cancer related excess mortality'. This outcome obviates the need to determine the cause of death. The basic principle of the method is an indirect standardisation of the total mortality in the breast cancer cases in the intervention group and in the control group. The reference population was assembled using data from the Swedish Cause of Death Register. The analysis indicated that screening reduced breast cancer related mortality among women aged 40 – 74 years by 24 per cent (Lar96).

Some of the Swedish RCTs used 'breast cancer present at death' as an endpoint, rather than 'breast cancer as underlying cause of death'. All Swedish RCTs were therefore analysed on the basis of one outcome or the other. Again, comparison of the results reveals little difference in the estimated effect (the relative risk reduction was 21 and 23 per cent respectively in women aged 40-74) (Nys95). This is an additional indication that the Swedish trials' results were not affected by serious bias.

Like other commentators (Mil01), the committee has considered whether the effect of radiotherapy on vascular mortality could have resulted in slippery-linkage bias. If such a bias could have been a problem at any stage, the committee doubts whether this was still the case after 1975, when the reviewed screening trials (except for the HIP trial) took place. Further consideration is given to the issues surrounding post-mastectomy radiotherapy and vascular disease in chapter 5. In the committee's view, it is not likely that radiotherapy could influence the findings in favour of screening.

4.2.3 Conclusion

The number of cases in which there is doubt whether breast cancer was the cause of death is much smaller than suggested in the Cochrane review. It is nevertheless true that breast-cancer mortality as an outcome measure in an RCT can be biased. As the Cochrane review correctly points out, blind (re)assessment of cause of death does not warrant protection against this possibility, since it cannot prevent slippery-linkage bias and the like. However, such forms of bias can be detected by the analysis of overall

mortality and breast cancer excess mortality. These analyses obviate the need to determine whether or not breast cancer was the cause of death in each individual case.

The committee cannot exclude the possibility that the findings of the reviewed RCTs may have been slightly biased, even after blind reassessment of the causes of death. However, there is no definite evidence of differential misclassification to an extent sufficient to have seriously biased the trial findings. The results of the breast cancer excess mortality analysis performed on the Swedish overview data also argue against serious bias. The committee therefore concludes that there is no reason to regard breast-cancer mortality as unreliable and biased in favour of screening. Nevertheless, because bias can never be excluded, total cancer mortality and overall mortality should be taken into account when interpreting the reduction in breast-cancer mortality.

Chapter 5

Overall mortality and total cancer mortality as an endpoint

5.1 Criticisms made in the Cochrane review

The reviewers regard overall mortality as the only suitable outcome measure. This is partly because they regard breast-cancer mortality as an unreliable outcome, and partly because they believe that the possibility of an *in*crease in deaths from other causes should be taken into account. Hence, an effect on overall mortality needs to be shown.

Such a screening-associated increase in mortality can result from post-mastectomy radiotherapy, it is suggested. In this context, the authors refer to the findings of a meta-analysis of radiotherapy trials carried out by the Early Breast Cancer Trialists' Collaborative Group (EBC00). From these findings, they cite that radiotherapy reduces annual breast-cancer mortalityby 13 per cent, but increases death from other causes by 21 per cent. This adverse effect is caused by an excess of cardiovascular deaths (EBC00).

The overview of the Swedish screening RCTs indicates a statistically non-significant rise in the intervention groups of death from all diseases of the circulatory system (Nys96). The Danes put the absence of statistical significance in this increase down to the short duration of the RCTs. The introduction of a national screening programme shortly after publication of the first favourable findings of the Two-County Trial (Tab85) made it very hard, they claim, to statistically distinguish any effect, whether positive or negative (due to contamination of the former control groups by the availability of screening for these women as well).

The Cochrane review concludes that screening does not reduce overall mortality (see also Annex E). The authors point out that the trials and the review are

underpowered for all-cause mortality; the confidence levels include both a plausible worthwile and a possible detrimental effect. The Cochrane review also suggests that screening is not associated with a decline in total cancer mortality (including breast cancer mortality). See also Annex D.

5.2 The committee's view

The committee regards breast-cancer mortality as an important endpoint (see chapter 4). The Danish authors' contention that overall mortality should be the primary endpoint in preference to breast-cancer mortality is not agreed with. It is nevertheless the committee's view that the effect of screening should be assessed not only on the basis of breast-cancer mortality, but in conjunction with total cancer mortality and overall mortality. These other outcome measures should be considered when interpreting the results of cancer-screening trials, as indeed they normally are (Duf01, EBC00, Höj99, Lar96, Nys96, Tab88, Tab89, Tab92). There are good reasons for this. First, such an approach facilitates the identification of any (direct or indirect) adverse effects that screening may have. Second, examination of all-cause mortality in combination with disease-specific mortality can reveal major threats to the validity of a RCT, such as flaws in randomisation and ascertainment of vital status (Ale89, Bla02).

The assertion that radiotherapy following mastectomy leads to increased vascular mortality and the expectation that radiotherapy even *in*creases overall mortality in screened women are derived from a meta-analysis (EBC00) of forty radiotherapy trials started between 1960 and 1990. Midway through this period, radiotherapy techniques were improved substantially (Kur00). For this reason, a number of large-scale trials were launched to determine the effect of modern locoregional radiotherapy, designed to minimise cardiopulmonary radiation exposure. The findings of these studies demonstrate that in the medium-term (median observation period: ten years) modern radiotherapy does not increase vascular morbidity and mortality, and that overall mortality actually declines (Höj99, Ove99). The ten-year survival rate was 45 per cent in the women who had received radiotherapy and 36 per cent in women who had not (Ove99). A meta-analysis of 18 recent studies reports very substantial benefits in irradiated patients with a decrease of overall mortality by 17 per cent: RR= 0.83 (0.74-0.94) (Whe00).

Longer follow-up is required to confirm that excess vascular mortality can be substantially reduced or even avoided by use of modern techniques. Nevertheless, we can already say that there appears to be no such excess mortality in the first twelve years. The effects have now been under observation for a period as long as that within which excess mortality was detected by the EBCTCG (Höj99). An update of the EBCTCG meta-analysis (EBC00) by Cuzick indicated that the trials started before

1975 had a very statistically significant increase in non-breast cancer deaths, while those started after 1975 did not. It was noteworthy that the majority of non-breast cancer deaths, although vascular in origin, were not cardiac related (Abr01). This delayed side-effect is thought to result from the radiation dose per fraction given prior to 1975, and the partial inclusion of the heart and the great vessels in the irradiated area. RCTs started after 1975 indicated that the beneficial effects of radiotherapy on local control and thus on the patient's chance of survival were no longer being all but cancelled out by excess vascular mortality (Abr01).

The overview of the Swedish screening RCTs (Nys96) did not reveal a rise in mortality from ischemic heart disease in the intervention groups: RR = 1.00 (0.95-1.05). Because a carotid artery sometimes lies within the irradiated area, it is also necessary to consider mortality from cerebrovascular diseases. The Swedish overview indicated a statistically non-significant 4 per cent rise in such mortality, relative to the control group: RR= 1.04 (0.97-1.12).

In any case, it does not strike the committee as likely that any harmful effects of radiotherapy would manifest themselves especially among screened women, as the Cochrane reviewers assume. Post-mastectomy radiotherapy is mainly used in cases where there is an increased risk of local breast cancer recurrence (tumour-positive axillary nodes, tumour diameter greater than 5 centimetres). Screening tends to detect less advanced cases of breast cancer, which can normally be treated without irradiating the lymph nodes in the neck (supraclavicular).

When interpreting the results of breast-cancer screening trials, it is also important to consider total cancer mortality. If screening has a positive effect on breast-cancer mortality, a reduction (albeit a much smaller drop) in all-cancer mortality is to be expected. The Danish authors are correct to state that none of the RCTs showed a statistically significant decline in all-cancer mortality (including breast-cancer mortality) (Annex D). However, the RCTs were underpowered for such a decline. The Swedish overview does show a statistically non-significant decrease of 2 per cent. It has been suggested that this decrease is too small, and that 30 per cent of the all-cancer mortality is attributable to breast cancer (Gia01, Gøt01). The failure to find evidence of a clear reduction in all cancer mortality is described as 'certainly a cause for concern'(Gia01). However, one should not overlook the fact that the proportion of all-cancer mortality accounted for by breast cancer will be much lower in a trial population than in the population at large: 11 per cent in the Swedish RCTs and 24 per cent among Dutch women (NCR00, Nys96). The reason being that women diagnosed with breast cancer at the start of the study are excluded from the trial population; the only deaths from breast cancer counted are those arising from breast cancers diagnosed during the trial time. The committee estimates that, if a 20 per cent reduction in

breast-cancer mortality were achieved in the relevant age group (forty to seventy-five), the resulting fall in all-cancer mortality in the Swedish RCTs should have been about 2 per cent. The non-significant decline referred to in the overview is consistent with this estimate, whereas the figures in the Cochrane review are not (see also Annex D). It should also be noted that the data on all-cancer mortality have not been broken down according to age. The committee feels it would be instructive to obtain insight into all cancer mortality among women over the age of fifty.

Finally, attention should be given to overall mortality as well. The Danish authors abandoned their assertion (Göt00) that mammographic screening causes six times as many deaths as it prevents. In the Cochrane review, it is stated that a screening-associated increase in mortality cannot be excluded. However, if available data from all eligible trials (excluding Edinburgh and New York) are considered then the relative risk reduction in overall mortality is one to three per cent for women aged 50 or over (see Annex E).

This is consistent with what might be expected. Among Swedish women aged between forty and seventy, breast cancer deaths constituted nearly 10 per cent of overall mortality in 1989 (Lar96). According to the overview, the corresponding percentage in the control groups of the Swedish RCTs was much lower, namely 3.4 per cent (Nys96). This was due to the fact that women diagnosed with breast cancer prior to the randomisation date were excluded from the trial. Hence, a 20 per cent reduction in breast-cancer mortality(or 25 per cent in women over the age of fifty) would be expected to bring down overall mortality by about 1 per cent.

An as yet unpublished update of the Swedish overview (median observation period nearly sixteen years) indicates that, among women aged between forty and seventy, overall mortality has declined by 2 per cent: RR= 0.98 (0.96-1.00) (Nys02).

Chapter 6

Overtreatment

6.1 Criticisms made in the Cochrane review

According to the Cochrane review, the estimated cumulative risk of a false positive result after 10 mammograms is 50 per cent. Screening also leads to more women being diagnosed with breast cancer and to increases both in the overall frequency of breast surgery and in more aggressive breast surgery, it is claimed. According to the Cochrane review, the rise in frequency of breast cancer cannot be dismissed as a temporary effect resulting from earlier diagnosis. Rather, it is a structural phenomenon attributable to the discovery of slow-growing cancers that would otherwise not have led to ill health. It is said that screening results in ductal carcinoma in situ (a precursor of breast cancer) being detected six times more often.

The authors claim that the resulting overtreatment involves 20 per cent more mastectomies and a 30 per cent increase in the overall number of breast operations (including breast-conserving surgery). It is also suggested that screening leads to a 25 to 40 per cent increase in the use of radiotherapy in the treatment of breast cancer.

6.2 The committee's view

When considering the benefit and risk of breast cancer screening, it is very important to take age into account. Women under the age of fifty are much more likely to get false positive screening results, in-situ cancer or interval cancer than older women (Fri86, Tab87). No such distinction is made in the Cochrane review. Notably, the

Cochrane Breast Cancer Group thought the conclusions on the use of more aggressive treatment were misleading, and they did not accept this section of the review (CBCG02).

6.2.1 *False-positive screening results*

The Cochrane review's alarming estimate that there is a 50 per cent cumulative risk of getting at least one false positive result after ten screening examinations is based upon American research (Elm98). This figure does not apply to the Netherlands. For one thing, it is based partly on data concerning women under the age of fifty, who are more liable to receive false positive results than older women. Furthermore, American doctors are more inclined than, for example, their Dutch counterparts to pursue a policy of 'defensive medicine', motivated by the fear of legal action. The American study (Elm98) indicates that four to eight screened women in a hundred receive a false positive result.

In the Netherlands, this percentage is much lower: 0.7 per cent where initial examinations are concerned and only half as many where subsequent examinations are concerned (Fra00). The rate of invasive interventions (biopsies) and the benign to malignant ratios at biopsy are also much lower in the Netherlands (and other European countries) than in the United States (see also Chapter 2).

6.2.2 *Overdiagnosis*

In essence, what the Cochrane review's authors say is correct. Overdiagnosis – the detection of breast cancer that may never have progressed to become symptomic during a woman's life– does indeed take place. That is an inherent problem with any screening strategy, and one that was anticipated prior to the decision to set up the national programme. Allowance for overdiagnosis was made when assessing the net benefits of the programme (Kon90, Maa87, WVC91). It is also the case that it is not possible to distinguish women whose in-situ cancer (ductal carcinoma in situ, a precursor of – invasive – breast cancer) or slow-growing (invasive) breast cancer will become aggressive due to further genetic changes from those whose condition will not deteriorate in this way.

It is indeed the case that the introduction of screening leads to an increase in the number of women being diagnosed with breast cancer. The increased frequency of breast cancer is associated not only with initial screening examinations (the prevalence pool), but also with subsequent examinations (lead time, discovery of ductal carcinomas in situ). On the other hand, the frequency of breast cancer is lower among women who are no longer being screened (because they decide not to participate, pass

the age of seventy-five or are unable to participate further due to termination of the programme). In fact, the frequency among such women is for a while lower than among women who have never been screened (Boe94). A few years after the HIP trial in New York had ended, there was no difference between the original intervention and control groups in terms of the (cumulative) number of breast cancer cases (Sha85, Sha90). This is not particularly instructive in the present context, however, since the mammographic technique used in the HIP trial involved a shorter lead time than that associated with modern techniques. More recent data is not available, since screening continued after completion of the other RCTs, and was widened to include the control group as well.

The Cochrane review makes no allowance for this compensation effect. If one looks only at the initial period following the introduction of screening, even if that is nearly ten years, it is inevitable that one will find an increase in the frequency of breast cancer and breast surgery. However, there is no evidence of any serious level of overdiagnosis in the longer term (Pee89).

Furthermore, the claims regarding overtreatment are based on incomplete data that do not distinguish between age groups under 50 and over 50 years or provide insight into the percentage distribution of the types of breast surgery. If such data had been presented, they would have shown that, particularly among screened women, there has been a relative decline in mastectomies and a relative increase in breast-saving surgery.

The level of overdiagnosis is much lower than suggested in the Cochrane review. Calculations made using MISCAN, a well-validated computer simulation model (Akk99, Kon90, Maa87), indicate that the Dutch screening programme increases the frequency of breast cancer diagnosis a few per cent (Boe94). The current percentage will be a little higher now that (as from 1998) women between the ages of seventy and seventy-five are also covered by the programme.

According to cancer registry data, 114 new cases of (invasive) breast cancer per 100,000 women were found in the Netherlands in 1990, and 127 in 1996. In the same period, the frequency of the diagnosis of in-situ cancer of the breast rose from five to eleven per 100,000 women. This rise is attributable mainly to the screening programme. Some 13 per cent of women between the ages of fifty and seventy with breast cancer detected by screening have in-situ cancer. The corresponding figure for women in the same age group with breast cancer detected during interscreening intervals or outside the context of screening is 3 to 6 per cent, depending on age (Fra01).

6.2.3 *Overtreatment*

In the Netherlands, among women between the ages of fifty and seventy with (invasive) breast cancer, the percentage receiving breast-conserving surgery rose from 32 per cent in 1990 to 42 per cent in 1996. In the same period, the percentage undergoing mastectomies fell from 59 per cent to 51 per cent (Fra01). In this country, therefore, there has been no increased use of aggressive treatment during the first seven years after the introduction of screening. In addition to greater use of breast-conserving surgery, the developmental phase of the programme (1990 to 1996) saw a decline in the use of both adjuvant therapy (hormonal therapy or chemotherapy) and radiotherapy. More than a third of treated women received adjuvant therapy – usually hormonal therapy where women over the age of fifty were concerned. If the treatment figures are broken down according to tumour stage (TNM status), no difference emerges between patients whose breast cancer was screen-detected and those whose breast cancer was discovered outside-screening (Fra01).

One of the reasons for more women being diagnosed with breast cancer is the raised detection rate of in-situ cancer, caused by the increasing use of mammography in routine diagnosis of breast disorders and the introduction of mammographic screening programmes. According to the Cochrane review, in-situ cancer is found six times as often where screening programmes are in place as where they are not. The reviewers state that in half of these cases the abnormality measures five centimetres or more, leading to amputation of at least one breast and often both breasts, although untreated in-situ cancer only develops into (invasive) breast cancer in 20 to 25 per cent of cases.

It is quite true that screening leads to a considerable rise in the detection of in situ-cancer, although the reported six-fold increase does include women under the age of fifty (Tab92). In the Netherlands between 1990 and 1996, 53 per cent of women between the ages of fifty and seventy in whom in-situ cancer was discovered by screening received breast-conserving surgery. Among women in the same age group who had not been screened, the figure was 42 per cent (Fra01).

The explanation for breast-conserving surgery not being performed more often in cases of in-situ cancer – which is, after all, a precursor of breast cancer – relates in part to difficulties inherent in the surgery. It is much harder for a surgeon to completely remove in-situ cancer (allowing a safe surrounding margin, Hol90) than to completely remove invasive breast cancer, because in-situ cancer is not generally palpable, but can be discerned only by mammography. (The dimension referred to should have been 2.5 centimetres, incidentally, not five). Failure to fully remove in-situ cancer appears to be the main cause of local recurrence following breast-conserving surgery (Hol98).

Such problems can be addressed by using magnification mammography, localization techniques, specimen radiography and other means to ensure complete excision of the mammographically detected abnormality, but this requires considerable interdisciplinary coordination (CBO88, GR87, GR98). In this field too, mammographic screening has been a strong force for quality improvement (Pag99), as indicated by the shift from mastectomy to breast-conserving surgery in the early years of the programme (Fra01).

In-situ cancer has long been something of a neglected field of medicine. On the assumption that in-situ cancer is typically multicentric (i.e. develops at various points within the breast), amputation of the breast has for many years been the standard therapy. However, the results of detailed pathologic and radiologic studies have revealed that the assumption of multicentric development is incorrect (Hol90).

Little is known with certainty about the natural history of in-situ cancer – partly because it cannot be studied without first removing breast tissue for establishing the diagnosis and, thereby, disturbing the natural history of these lesions. Nevertheless, the committee does *not* believe that in-situ cancer should be regarded as a biologically benign condition, which if left alone would not develop into (invasive) breast cancer in 70 to 75 per cent of cases. Histopatholopically, it is possible to distinguish various forms of in-situ cancer, which can be divided into three malignancy grades. Long-term studies on the progression to invasive breast cancer of in-situ lesions originally misdiagnosed as benign indicate a 25 to 50 per cent progession rate at 30 years of follow-up for low-grade lesions (Eus94, Pag95). Although little is known about the natural history of high-grade in-situ cancer, research involving a small number of patients has suggested that there may be a 50 to 75 per cent chance of progression in four to nine years (Eus94, Eva01). The findings of clinical follow-up studies of women after breast-conserving surgery indicate that the risk of recurrence is greater with the most malignant forms of in-situ cancer (comedonecrosis) than with less malignant forms. Roughly 65 per cent of screen-detected lesions are high grade, and 13 per cent a low-grade form (Eva01). In other words, it is not the case that screening mainly detects more indolent forms of in-situ cancer. More malignant forms of in-situ cancer (comedonecrosis) are often easier to detect mammographically because of the characteristic micro-calcifications. Further research is required to provide greater insight into the balance between overtreatment and the prevention of (invasive) breast cancer.

6.2.4 *Concluding remarks*

On the basis of detailed data from three Swedish and two Dutch screening projects, the MISCAN model has been used to make careful long-term estimates of the benefits and

harms of breast-cancer screening in the Netherlands (Kon90, Maa87). Optimum performance of the screening programme and the diagnostic work-up of positive screenees can be expected to lead to a situation as outlined below.

Roughly 27 per cent of the women in whom breast cancer is detected by screening will survive the disease as a result of early detection. For the remaining 73 per cent, the chances of survival will be unaffected, despite detection of the disease a few years earlier than otherwise have been the case. Some 53 per cent of the women would have been treated in time to save their lives even if their cancer had not been detected by screening. However, some of these women now have the option of breast-saving surgery, which they would not have had without early detection. A further 13 per cent will die despite the earlier detection of their cancer. Finally, there is a group of up to 7 per cent who would never have known about their cancer if it were not for screening, since they would have died of other causes before clinical manifestation of the disease (Kon90, LETB95, Maa01).

Screening enables breast cancer to be detected earlier, when the tumour is smaller and metastasis to the lymph nodes is less likely to have occurred. This means that ‘aggressive’ forms of diagnostic and therapeutic intervention (e.g. excision biopsy, axillary sampling or dissection and mastectomy) are necessary less often. In recent decades, screening has accelerated the development of less mutilating forms of intervention, thereby reducing suffering and the cosmetic implications of the disease. The Cochrane review ignores these facts, choosing only to look backwards.

Chapter 7

Observational studies

The Cochrane review rightly focuses primarily on RCTs. Such experimental studies provide the best evidence to resolve scientific issues. However, it is also worth considering observational research: case-control studies and cohort studies. Although the findings of such non-experimental research carry less weight, they can be useful. This chapter summarizes the results of a number of observational studies in this field.

7.1 Case-control studies

The results of case-control studies carry less weight than RCT findings, because of the potential for bias. Nevertheless, the results of well-designed case-control studies are fairly consistent with those of RCTs (Ben00, Con00, Gil95). This is true in general and of research into mammographic screening in particular (Con00, Dem98, Ker95). The case-control studies in Nijmegen, Utrecht, Florence, Malmö, the United Kingdom and the United States all point to a clear reduction in breast-cancer mortality among women over the age of fifty (Col84, Col92, Gul91, Jan90, Mos92, Pal89, Tho94, Ver84).

7.2 Effect evaluation

The purpose of observational research into the impact of service screening on long-term breast-cancer mortality is not to prove the effectiveness of screening. Such research is not capable of providing proof, because there is no way of being certain

that any observed change in mortality is not attributable to some other factor. Only if the screening programme is introduced in the context of an RCT, as was done in Finland (Hak97), the effects of other factors are controlled for. Observational research in the form of ongoing evaluation (monitoring) studies of service screening programmes can nevertheless provide additional information regarding the effectiveness of screening.

The authors of the Cochrane review say that their work was prompted partly by such a study whose results were published in a Swedish journal (*Läkartidningen 1999*; 96: 904-13). This report indicated that breast-cancer mortality should have fallen by 11 per cent following the introduction of service screening in Sweden between 1986 and 1997, but had in fact declined by barely 1 per cent.

From the ensuing correspondence letters in English-language journals, it is apparent that a number of important criticisms have been levelled at the study into breast-cancer mortality in Sweden, which have not been adequately rebutted (May99, Nys00, Ros99, Sjö99). The study lacked sufficient statistical power to distinguish the effect of screening, and it was wrongly assumed that a linear decline in breast-cancer mortality could be expected right from 1986. Furthermore, no allowance was apparently made for the rising incidence of breast cancer in the period before 1986, for the fact that the programmes started at different times in the 17 Swedish counties or for the breast-cancer mortality among women diagnosed with breast cancer before the screening programmes started.

In Sweden, contrary to the reviewers' contention, breast-cancer mortality barely changed between 1950 and 1985 (Bor95). The introduction of Sweden's breast-cancer screening programme was completed in 1997 (Ols00). As in Finland, a positive effect on breast-cancer mortality was subsequently observed in a number of studies (Hak97, Jon01, Tab01). The extent of this effect was broadly consistent with what one might expect from the screening trials' findings.

In the United Kingdom, breast-cancer mortality suddenly started to fall markedly after 1987, particularly among middle-aged women, following a period in which it had risen sharply (Pet00, Qui95). Since screening was introduced to the UK between 1988 and 1995, the fall cannot be attributable only to screening. A decline in national breast-cancer mortality due to screening will not be seen until several years after introduction, because mortality rates are affected for some while by women who were diagnosed with breast cancer before the screening programme started. The sudden fall seen in the UK is therefore more likely to have been due to improvements in treatment – the more widespread use since the 1980s of adjuvant therapy with tamoxifen and chemotherapy, together with improvements in radiotherapy and surgery (Pet00). From 1987, breast-cancer mortality also began falling in the United States, where there is no organised screening programme but a lot of screening nevertheless takes place. British

researchers have since reported a small but unconvincing decline in breast-cancer mortality attributable to screening (Bla00).

After decades of stability, breast-cancer mortality in the Netherlands has been falling gradually since 1991. In 1990, it was estimated that the effect of screening on overall breast-cancer mortality (i.e. among all age groups) should reach half of its maximum by about 1999, the most recent year for which mortality data are presently available. (The maximum forecast reduction being 16 per cent, expected by 2015; see chapter 2.) Using the computer simulation model MISCAN, forecasts of breast-cancer mortality in the Netherlands were made in 1990, first assuming the introduction of a screening programme, then assuming no programme (Akk00). Both in the years before the introduction of screening, and between 1989 and 1998, observed mortality was consistent with the figures forecast assuming the existence of a programme. In 1997 and 1998, the observed mortality was significant lower than in 1986-1988, before screening started. Comparing age-standardised mortality rates, with and without screening, observed and predicted by MISCAN, it was found that the mortality rate was significantly lower than the expected mortality rate without screening for the 60-69 age group in 1997, and for the 50-59 and 60-69 age groups in 1998 (Fra00). In 1999, the decline continued in the original screening target group (Fra01). With a view to determining the extent to which the observed reduction in breast-cancer mortality is due to screening, or to improved treatment, the LETB is to link individual cause-of-death data with data from the screening programme and the cancer registry (Fra00).

Such research requires data relating to women who have *not* participated in the screening programme. This presents an as yet unresolved problem, since women who have declined to be screened are not able to give their consent for the release of the relevant data from the cancer registry.

Chapter 8

Answers to questions posed by the minister

The minister asked: Do the findings of the Cochrane review nullity the scientific basis for breast-cancer screening for women over the age of fifty? In this context, it is also important to consider whether breast-cancer mortality is a valid outcome.

The committee regards breast-cancer mortality an important outcome measure of trials of breast cancer screening, provided that such mortality is reliably determined. It does seem to the committee that the use of disease-specific mortality as the primary endpoint renders screening trials subject to bias. Whether that was actually the case or not, the committee believes that breast-cancer mortality should be examined in conjunction with all-cancer mortality and overall mortality when interpreting the results of screening trials. Appendices C, D and E provide an illustrative overview, based on the findings of the Cochrane review, reworked by the Dutch Cochrane Centre in Amsterdam.

All the reviewed RCTs show a reduction in breast-cancer mortality among women over the age of fifty, although it is not statistically significant in all cases. If data from all eligible RCTs (except for Edinburgh and New York) are considered then the risk of dying from breast cancer is reduced by 20 per cent among women between the ages of forty and seventy-five (Ols01b). For women over the age of fifty, the reduction is 28 per cent after seven years and 25 per cent after thirteen years (including New York; see Annex C).

On the basis of findings from Kopparberg, Östergötland and Malmö, it was forecast that the national screening programme in the Netherlands would bring down

breast-cancer mortality by 33 per cent among women between the ages of fifty and seventy (Kon90).

On 15 February 2002, the findings of an update of the Swedish overview were presented in Stockholm. The updated data cover an observation period with a median duration of nearly sixteen years – seven years longer than the period covered by the original data. The conclusion is that breast-cancer mortality among women aged forty to seventy has been cut by 21 per cent: RR = 0.79 (0.70-0.98). Overall mortality has fallen by 2 per cent: RR = 0.98 (0.96-1.00) (Nys02).

The Cochrane review rightly focuses primarily on RCTs. However, observational research into the value of mammographic screening has also been conducted. The results of this research reinforce the conclusions of the experimental research regarding the benefits of screening for women over the age of fifty.

In the committee's view, the currently available reliable evidence does show a survival benefit of breast-cancer screening for women over the age of fifty. The arguments presented in the Cochrane review are not considered convincing to refute this evidence. However, the committee does not discount the possibility that new evaluations will ultimately lead to the conclusion that the effect of screening is not as great as anticipated in 1990.

References

Abr01 Abrams JS, Eifel P. Monograph overview. Monogr Natl Cancer Inst 2001; 30: 1-4.

Ald67 Alderson M, Meade T. Accuracy of diagnosis on death certificates compared with that in hospital records. Br J Prev Soc Med 1967; 21: 22-9.

Akk99 van den Akker-van Marle E, de Koning H, Boer R, *et al*. 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: 30-4.

Alb00 Albertsen PC, Walters S, Hanley JA. A comparison of cause of death determination in men previously diagnosed with prostate cancer who died in 1985 or 1995. J Urol 2000; 163: 519-23.

Ale89 Alexander F, Roberts MM, Lutz W, *et al*. Randomisation by cluster and the problem of social class. J Epidemiol Community Health 1989; 43: 29-36.

Ale99 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.

And88 Andersson I, Aspegren K, Janzon L, *et al*. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. Br Med J 1988; 297: 943-8.

Bai97 Bailar III JC, MacMahon B. Randomisation in the Canadian National Breast Screening Study: a review for evidence of subversion. Can Med Assoc J 1997; 156: 193-9.

Ben00 Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000; 342: 1878-86.

Bla00 Blanks RG, Moss SM, McGahan CE, *et al*. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990-8: comparison of observed with predicted mortality. Br Med J 2000; 665-9.

Bla02 Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. J Natl Cancer Inst 2002; 94: 167-73.

Bju97 Bjurstam N, Björneld L, Duffy SW, *et al*. The Gothenburg breast screening trial: preliminary results on breast-cancer mortalityfor women aged 39-49. Monogr Natl Cancer Inst 1997; 22: 53-5.

Boe94 Boer R, Warmerdam P, de Koning H, *et al*. Extra incidence caused by mammographic screening. Lancet 1994; 343: 979.

Bor95 Bornefalk A, Persson I, Bergström R. Trends in breast-cancer mortality among Swedish women 1953-92: analysis by age, period and birth cohort. Br J Cancer 1995; 72: 493-7.

Boy97 Boyd NF. The review of randomisation in the Canadian National Breast Screening Study. Can Med Assoc J 1997; 156: 207-9.

Bre95 Brekelmans CT, Peeters PM, Deurenberg JJ, Collette HJA. Survival in interval breast cancer in the DOM Screening Programme. Eur J Cancer 1995; 31A: 1830-5.

Bri84 Brinkley D, Haybrittle JL, Alderson MR. Death certification in cancer of the breast. Br Med J 1984; 289: 465-7.

Bro93 Brown BW, Brauner C, Minnotte MC. Noncancer deaths in white adult cancer patiens. J Natl Cancer Inst 1993; 85: 979-87.

Bur96 Burrell HC, Sibbering DM, Wilson ARM, *et al*. Screening interval breast cancers: mammographic features and prognostic factors. Radiology 1996; 199: 811-7.

Cat00 Cates C, Senn S. Screening mammography re-evaluated. Lancet 2000; 355: 750.

CBCG02 The Cochrane Breast Cancer Group, Gøtzsche P, Horton R, red. Screening mammography: setting the record straight. Lancet 2002; 359: 439-42.

CETS93 Conseil d' évaluation des technologies de la santé du Québec. Screening for breast cancer in women aged 40-49 years. Montreal: CETS, 1993.

Cha91 Chamberlain J, Coleman D, Ellman R, *et al*. Verification of the cause of death in the trial of early detection of breast cancer. Br J Cancer 1991; 64: 1151-6.

Col84 Collette HJA, Day NE, Rombach JJ, *et al*. Evaluation of screening for breast cancer in a non-randomised study (the DOM Project) by means of a case-control study. Lancet 1984; i: 1224-6.

Col92 Collette HJA, de Waard F, Rombach JJ, *et al*. Further evidence of benefits of a (non-randomised) breast-cancer screening programme: the DOM project. J Epidemiol Community Health 1992; 46: 382-6.

Col98 Collins S, Woodman CBJ, Threlfall A, Prior P. Survival rates from interval cancer in NHS breast screening programme. BMJ 1998; 316: 832-3.

Con00 Concato J, Shah N, Horwitz RI. Randomized controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000; 342: 1887-92.

Cox97 Cox B. Variation in the effectiveness of breast screening by year of follow-up. Monogr Natl Cancer Inst 1997; 22: 69-72.

Cuc91 Cuckle H. Breast-cancer screening by mammography: an overview. Clin Radiol 1991; 43: 77-80.

Dem98 Demissie K, Mills OF, Rhoads GG. Empirical comparison of the results of randomized controlled trials and case control studies in evaluating the effectiveness of screening mammography. J Clin Epidemiol 1998; 51: 81-91.

Duf01 Duffy SW. Interpretation of the breast screening trials: a commentary on the recent paper by Gøtzsche and Olsen. Breast 2001; 10: 209-12.

Dijc96 van Dijck JAAM. The efficacy of mammographic screening for breast cancer in elderly women. (Dissertation). Nijmegen: Katholieke Universiteit Nijmegen, 1996.

EBC00 Early Breast cancer Trialists' Collaborative Group. Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: An overview of the randomised trials. Lancet 2000; 355: 1757-70.

Ede99 Ederer F, Geisser MS, Mongin SJ, *et al*. Colorectal cancer deaths as determined by expert committee and from death certificate: a comparison. The Minnesota Study. J Clin Epidemiol 1999; 52: 447-52.

Elm98 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.

Eus94 Eusebi V, Feudale E, Foschini MP, *et al*. Longterm follow-up of in situ carcinona of the breast. Semin Diagn Pathol 1994; 11: 223-35.

Eva01 Evans AJ, Pinder SE, Ellis IO, Wilson ARM. Screen detected ductal carcinoma in situ (DCIS): overdiagnosis or an obligate precursor of invasive disease? J Med Screen 2001; 8: 149-51.

Fle93 Fletcher SW, Black W, Harris R, *et al*. Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst 1993; 85: 1644-56.

Fow01 Fowble B, Hanlon A, Freedman G, *et al.* Second cancers after conservative surgery and radiation for stages I-II breast cancer. Int J Radiat Oncol Biol Phys 2001; 51: 679-90.

Fra00 Fracheboud J, Groenewoud JH, Boer R, *et al*. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland. Het achtste evaluatierapport. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit Rotterdam, 2000.

Fra01 Fracheboud J, Otto SJ, Groenewoud JH, *et al.* Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland. Het negende evaluatierapport. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit Rotterdam; 2001.

Fri86 Frisell J, Glas U, Hellström L, *et al*. Randomized mammographic screening for breast cancer in Stockholm. Breast cancer Res Treat 1986; 8: 45-54.

Fri91 Frisell J, Eklund G, Hellström L, *et al.* Randomized study of mammography screening – preliminary report on mortality in the Stockholm trial. Breast cancer Res Treat 1991; 18: 49-56.

Gar96 Garne JP, Aspegren K, Balldin G. Breast cancer as cause of death. A study of the validity of the officially registered cause of death in 2631 breast cancer patients dying in Malmö, Sweden 1964-1992. Acta Oncol 1996; 35: 671-5.

Gia01 Giard RWM, Bonneux LGA. Borstkankerscreening onvoldoende effectief. Ned Tijdschr Geneeskd 2001; 145: 2205-8.

Gil95 Gill TM, Horwitz RI. Evaluating the efficacy of cancer screening: clinical distinctions and case-control studies. J Clin Epidemiol 1995; 48: 281-92.

Gøt00 Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet 2000; 355: 129-34.

Gøt01 Gøtzsche PC. Screening for breast cancer with mammography. Lancet 2001; 358: 2167-8.

GR87 Gezondheidsraad. De vroege opsporing van borstkanker. Den Haag: Gezondheidsraad, 1987; publicatie nr 1987/09.

GR98 Gezondheidsraad. Wet bevolkingsonderzoek: landelijke borstkankerscreening. Rijswijk: Gezondheidsraad, 1998; publicatie nr 1998/03WBO.

Gul91 Gullberg B, Andersson I, Janzon, Ranstam J. Screening mammography. Lancet 1991; 337: 244.

Hak97 Hakama M, Pukkala E, Heikkilä M, *et al*. Effectiveness of the public health policy for breast-cancer screening in Finland : population based cohort study. Br Med J 1997; 314: 864-7.

Hay00 Hayes C, Fitzpatrick P, Daly L, Buttimer J. Screening mammography re-evaluated. Lancet 2000; 355: 749.

Höj99 Höjris I, Overgaard M, Christensen JJ, *et al*. Morbidity and mortality of ischaemic heart disease in high-risk breast cancer patiens after adjuvant postmastectomy systemic treatment with or without radiotherapy: analysis of DBCG82b and 82c randomised trials. Lancet 1999; 354: 1425-30.

Hol86 Holmberg LH, Tabár L, Adami HO, Bergström R. Survival in breast cancer diagnosed between mammographic screening examinations. Lancet 1986; ii: 27-30.

Hol90 Holland R, Hendriks JHCL, Verbeek ALM. Extent, distribution, and mammographic/histological correlations of breast ductal carcinoma in situ. Lancet 1990; 335: 519-22

Hol98 Holland PA, Gandhi, Knox WF, *et al*. The importance of complete excision in the prevention of local recurrence of ductal carcinoma in situ. Br J Cancer 1998; 77: 110-4.

Hol00 Holland R. Met vroege ontdekking zijn wij er nog niet. Inaugural spreech, in Dutch. Nijmegen: KUN, 2000

Ioa98 Ioannidis JPA,Cappellari JC, Lau J. Issues in comparisons between meta-analyses and large trials. JAMA 1998; 279: 1089-93

Jon01 Jonsson H, Nyström L, Törnberg S, *et al.* Service screening with mammography of women aged 50-69 years in Sweden: effects on mortality from breast cancer. J. Med Screen 2001; 8: 152-60

Juf02 Juffs HG, Tannock IF. Screening trials are even more difficult than we thought they were. J Natl Cancer Inst 2002; 94; 156-7.

Ker95 Kerlikowske K, Grady D, Rubin SM, *et al*. Efficacy of screening mammography: a meta-analysis. JAMA 1995; 273: 149-54.

Kir85 Kircher T, Nelson J, Burdo H. The autopsy as a measure of accuracy of the death cerificate. N Engl J Med 1985; 313: 1263 - 9.

Kja01 Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001; 135: 982-9

Kon90 de Koning HJ, Boer R, van der Maas PJ, *et al*. Effectiviteit van bevolkingsonderzoek naar borstkanker; sterftereductie in binnen- en buitenland. Ned Tijdschr Geneeskd 1990; 134: 2240-5.

Kon90 de Koning HJ, van Ineveld BM, van Oortmarssen GJ, *et al*. De kosten en effecten van bevolkingsonderzoek naar borstkanker. Eindrapport. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit, 1990.

Kon95 de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PMM, *et al.* Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. J Natl Cancer Inst 1995; 87: 1217-23.

Kon00 de Koning HJ. Assessment of nationwide cancer-screening programmes. Lancet 2000; 355: 80-1.

Kop97 Kopans DB. Canadian National Breast Screening Study. Lancet 1997; 350: 810.

Kur00 Kurtz JM. Radiotherapy for early breast cancer: was a comprehensive overview of trials needed? Lancet 2000; 355: 1739-440.

Lar96 Larsson L-G, Nyström L, Wall S, *et al.* The Swedish randomised mammography screening trials: analysis of their effect on the breast cancer related excess mortality. J Med Screen 1996; 3: 129-32.

Law00 Law M, Hackshaw A, Wald N. Screening mammography re-evaluated. Lancet 2000; 355: 749-50.

LETB95 Landelijk Evaluatie Team voor bevolkingsonderzoek naar Borstkanker. Landelijke evaluatie van bevolkignsonderzoek naar borstkanker in Nederland. IV. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit Rotterdam, 1995

Maa87 van der Maas PJ, van Ineveld BM, van Oortmarssen GJ, *et al.* De kosten en effecten van bevolkingsonderzoek op borstkanker. Interimrapport 1987. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit, 1987.

Maa00 van der Maas PJ. Bevolkingsonderzoek naar borstkanker: een tussenbalans. Ned Tijdschr Geneeskd 2000; 144: 1096-9.

Maa01 van der Maas PJ. Breast-cancer screeningprogramme in the Netherlands: an interim review. The Breast 2001; 10: 12-4.

Mat00 Matsuyama Y, Tominaga T, Nomura Y, *et al*. Second cancers after adjuvant tamoxifen therapy for breast cancer in Japan. Ann Oncol 2000; 11: 1537-43.

May99 Mayor S. Swedish study questions mammography screening programmes. Br Med J 1999; 318: 621.

Met93 Mettlin CJ, Smart CR. The Canadian National Breast Screening Study. Cancer 1993; 72: 1461-5.

Mie02 Miettinen OM, Henschke CI, Pasmantier MW, *et al*. Mammographic screening: no reliable supporting evidence? Lancet 2002; 359: 404-6.

Mil92 Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: I and II. Can Med Assoc J 1992; 147: 1459-88.

Mil00 Miller AB, To T, Baines CJ, Wall C. 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.

Mil01 Miller AB. Screening for breast cancer with mammography. Lancet 2001; 358: 2164.

Moh98 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.

Mos92 Moss SM, Summerley ME, Thomas BT, *et al*. A case-control evaluation of the effect of breast-cancer screening in the United Kingdom trial of early detection of breast cancer. J Epidemiol Community Health 1992; 46: 362-4.

Nar97 Narod SA. On being the right size: a reappraisal of mammography trials in Canada and Sweden. Lancet 1997; 349: 1846.

NCR00 Visser O, Coebergh JWW, Schouten LJ, van Dijck JAAM (editors). Incidence of cancer in the Netherlands 1996. Eight report of the Netherlands Cancer Registry. Utrecht: Vereniging van Integrale Kankercentra, 2000

New00 Newschaffer CJ, Otani K, McDonald MK, Penberthy LT. Causes of death in elderly prostate cancer patients and in a comparison nonprostate cancer cohort. J Natl Cancer Inst 2000; 92: 613-21.

Nix00 Nixon R. Some random-effect models for the analysis of matched-cluster randomised trials: application to the Swedish two-country trial of breast cancer screening. J Epidemiol Biostat 2000; 5: 349-58.

Nom99 Nomura U, Tsutsui S, Murakami S. Takenaka Y. Prognostic impact of second cancer on the survival of early breast cancer patients. Int J Oncol 1999; 14: 1103-9.

Nys93 Nyström L, Rutqvist LE, Wall S, Lindgren A, *et al*. Breast-cancer screening with mammography: overview of Swedish randomised trials. Lancet 1993; 341: 973-8.

Nys95 Nyström L, Larsson L-G, Rutqvist LE, *et al.* Determination of cause of death among breast cancer cases in the Swedish randomized mammography screening trials. A comparison between official statistics and validation by an endpoint committee. Acta Oncol 1995; 34: 145-52.

Nys96 Nyström L, Larsson L-G. Wall S, Rutqvist LE, *et al*. An overview of the Swedish randomised mammography trials: total mortality patterns and the representivity of the study cohorts. J Med Screening 1996; 3: 85-7.

Nys00 Nyström L. Screening mammography re-evaluated. Lancet 2000; 355: 748-9

Nys02 Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 2002; 359: 909-19.

Obe00 Obedian E, Fischer DB, Haffty BG. Second malignancies after treatment of early-stage breast cancer: lumpectomy and radiation therapy versus mastectomy. J Clin Oncol 2000; 18: 2406-12.

Ols00 Olsson S, Andersson I, Karlberg I, *et al*. Implementation of service screening with mammography in Sweden: from pilot study to nationwide programme. J Med Sreen 2000; 7: 14-8.

Ols01 Olsen O, Gøtzsche PC. Systematic review of screening for breast cancer with mammography. London: Lancet Publishing Group, 2001: 1-46 (available at http: //image.thelancet. com/lancet/extra/fullreport.pdf).

Ols01a Olsen O, Gøtzsche PC. Cochrane review on screening for breast cancer with mammography. Lancet 2001; 358: 1340-2.

Ols01b Olsen O, Gøtzsche PC. Screening for breast cancer with mammography (Cochrane review). In: The Cochrane Library, Issue 4, 2001. Oxford: Update Software (www.cochranelibrary.net).

Ove99 Overgaard M, Jensen M-B, Overgaard J, *et al*. Postoperative radiotherapy in high-risk postmenopausal breast cancer patients given adjuvant tamoxifen: Danish Breast Cancer Cooperative Group DBCG82c randomised trial. Lancet 1999; 353: 1641-8.

Pac97 Paci E, Alexander FE. Study design of randomized controlled clinical trials of breast cancer screening. J Natl Cancer Inst 1997; 22: 21-5.

Pag95 Page DL, Dupont WD, Rogers LW, *et al*. Continued local recurrence of carcinoma 15-25 years after a diagnosis of low grade ductal carcinoma in situ of the breast treated only by biopsy. Cancer 1995; 76: 1197-200.

Pag99 Page DL, Simpson JF. Ductal carcinoma in situ – The focus for prevention, screening, and breast conservation in breast cancer. N Engl J Med 1999; 340: 1499-1500.

Pal89 Palli D, Del Turco MR, Buiatti E, *et al*. A case-control study of the effectiveness of a non-randomized breast-cancer screening program in Florence (Italy). Int J Cancer 1989; 43: 226-30.

Pee89 Peeters PHM, Verbeek ALM, Straatman H, *et al*. Evaluation of overdiagnosis of breast cancer in screening with mammography: results of the Nijmegen Programme. Int J Epidemiol 1989; 18: 295-9.

Pen01 Penson DF, Albertsen PC, Nelson PC, *et al*. Determining cause of death in prostate cancer: are death certificates valid? J Natl Cancer Inst 2001; 93: 1822-3.

Pet00 Peto R, Boreham J, Clarke M, *et al*. UK and USA breast cancer deaths down 25% in year 2000 at ages 20 - 69 years. Lancet 2000; 355: 1822.

Qui95 Quinn M, Allen E. Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. Br Med J 1995; 311: 1391-5.

Ros99 Rosén M, Rehnqvist N. No reed to reconsider breast screening programme on basis of results from defective study. Br Med J 1999; 318: 809-10.

Rub00 Rubino C, de Vathaire F, Diallo I, *et al.* Increased risk of second cancer following breast cancer: role of the initial treatment.Breast cancer Res Treat 2000; 61: 183-95.

Sat98 Satariano WA, Ragland KE, *et al*. Cause of death in men diagnosed with prostate carcinoma. Cancer 1998; 83: 1180-8.

Sch90 Schmidt JG. The epidemiology of mass breast-cancer screening– a plea for a valid measure of benefit. J Clin Epidemiol 1990; 43: 215-25.

Sch95 Schulz KF, Chalmers I, Hayes RJ, *et al*. Empiric evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273: 408-12.

Sch96 Schroen AA, Wobbes T, van der Sluis RF. Interval carcinomas of the breast: a group with intermediate outcome J Surg Oncol 1996; 63: 141-4.

Sch99 Scholten RJPM, Assendelft WJJ, Kostense PJ, Bouter LM. De praktijk van systematische reviews. V. Heterogeniteit tussen onderzoeken en subgroepsanalysen. Ned Tijdschr Geneeskd 1999; 143: 843-8.

Sen01 Senn S. Screening for breast cancer with mammography. Lancet 2001; 358: 2165.

Sha82 Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten to fourteen-year effect of screening on breast cancer mortality. J Natl Cancer Inst 1982; 69: 349-55.

Sha85 Shapiro S, Venet W, Strax Ph, *et al*. Selection, follow-up, and analysis in the Health Insurance Plan Study. Monogr Natl Cancer Inst 1985; 67: 65-74.

Sha90 Shapiro S. A dissent from Dr Schmidt's appraisal of evidence on breast cancer screening. J Clin Epidemiol 1990; 43: 227-34.

Sha97 Shapiro S. Periodic screening for breast cancer: the HIP randomized controlled trial. Monogr Natl Cancer Inst 1997; 22: 27-30.

Sjö99 Sjönell G, Ståhle L. Scientific foundation of mammographic screening is based on inconlusive research in Sweden. Br Med J 1999; 319: 55.

Tab85 Tabár 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; i: 829-32.

Tab87 Tabár L, Fagerberg G, Day NE, *et al*. What is the optimum interval between mammographic screening examinations? – an analysis based on the latest results of the Swedish Two-County Breast Cancer Screening Trial. Br J Cancer 1987; 55: 547-51.

Tab88 Tabár 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, editors. Screening for breast cancer. Toronto: Hans Huber, 1988: 39-44.

Tab89 Tabár L, Fagerberg G, Duffy SW, *et al*. The Swedish Two-county trial of mammography screening for breast cancer: recent results and calculation of benefit. J Epidemiol Community Health 1989; 43: 107-14.

Tab92 Tabár L, Fagerberg G, Duffy SW, *et al*. Update of the Swedish Two-county program of mammographic screening for breast cancer. Radiol Clin North Am 1992; 30: 187-210.

Tab99 Tabár L, Duffy SW, Vitak B. The natural history of breast carcinoma: what have we learned from screening? Cancer 1999; 86: 449-62.

Tab01 Tabár L, Vitak B, Chen H-HT, Yen M-F, *et al*. Beyond randomized controlled trials. Organized mammographic screening substantially reduces breast carcinoma mortality. Cancer 2001; 91: 1724-31.

Tar95 Tarone RE. The excess of patients with advanced breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. Cancer 1995; 75: 997-1003.

Tho94 Thompson RS, Barlow WE, Taplin SH, *et al*. A population-based case-cohort evaluation of the efficacy of mammographic screening for breast cancer. Am J Epidemiol 1994; 140: 889-901.

Ver84 Verbeek ALM, Hendriks JHCL, Holland R, *et al*. Reduction of breast-cancer mortality through mass screening with modern mammography. Lancet 1984; i: 1222-4.

Wal93 Wald NJ, Chamberlain J, Hackshaw A, on behalf of the Evaluation committee. Consensus statement: report of the European Society for Mastology Breast Cancer Screening Evaluation Commitee (1993). Breast 1993; 2: 209-16.

Whe00 Whelan TJ, Julian J, Wright J, *et al*. Does locoregional radiation therapy improve survival in breast cancer? A meta-analysis. J Clin Oncol 2000; 18: 1220-9.

WVC91 Notitie over het bevolkingsonderzoek naar borstkanker. Brief van de Staatssecretaris van Welzijn, Volksgezondheid en Cultuur aan de Voorzitter van de Tweede Kamer der Staten-Generaal. Tweede Kamer, vergaderjaar 1990-1991, 18 833, nr. 7. ‘s-Gravenhage: Sdu Uitgeverij, 1991.

A The minister’s request for advice

B The committee

C Breast cancer mortality

D Total cancer mortality

E Overall mortality

Annexes

Annex **A**

Request for advice

On 12 February 2002, the Minister of Health, Welfare and Sport wrote as follows to the President of the Health Council (letter reference GZB/GZ 2.250.678):

Following the recent appearance of articles in The Lancet and the *Nederlands Tijdschrift voor Geneeskunde* questioning the effectiveness of breast cancer screening, I hereby request that you prepare an advisory report for me regarding this issue within the Dutch context.

To be specific, I wish to know whether the findings of the meta-analysis published in the *Lancet* in October 2001 nullify the scientific basis for breast-cancer screeningas it is organized in the Netherlands (with women over the age of fifty as the target group and annual evaluation).

(signed)
Dr E Borst-Eilers

Annex **B**

The committee

- Professor JA Knottnerus, *Chair*
  President of the Health Council; The Hague
- Professor GMM Bartelink
  Radiotherapist; Antoni van Leeuwenhoek Hospital, Amsterdam, professor of radiotherapy, University Medical Centre, Amsterdam
- Professor Y van der Graaf
  Clinical epidemiologist; University Medical Centre, Utrecht
- Professor JGM Klijn
  Oncologist; Erasmus University Medical Centre and Daniël den Hoed Clinic, Rotterdam
- Professor FE van Leeuwen
  Epidemiologist; Dutch Cancer Institute, Amsterdam, professor of cancer epidemiology, Free University Medical Centre, Amsterdam
- Dr RJPM Scholten
  Clinical epidemiologist;, Dutch Cochrane Centre, University Medical Centre Amsterdam
- Dr WA van Veen, *secretary*
  Health Council, The Hague

The committee also met and consulted with the following experts:

- Dr RWM Giard, Pathologist and clinical epidemiologist; MRCZ, Clara site, Rotterdam

- Professor R Holland, Pathologist; National Reference Centre for Breast cancer Screening, St Radboud University Medical Centre, Nijmegen
- Dr HJ de Koning, Clinical epidemiologist; National Evaluation Team for Breast cancer Screening, Erasmus University Medical Centre, Rotterdam
- Professor ALM Verbeek, Clinical epidemiologist; St Radboud University Medical Centre, Nijmegen

Annex **C**

Breast cancer mortality

See tables on the following pages.

Table C1 Breast cancer deaths (obs), number of women (N) and relative risk (RR), with 95 per cent confidence interval (95% CI) for death from breast cancer in the intervention group (i.e. women invited to mammographic screening), compared with the control group (i.e. those not invited). Women aged fifty (Malmö fifty-five) or older at the start of the trial. Observation period: seven years.

trial	intervention group		control group		RR	95% CI
	Obs	N	Obs	N		
Malmö	35	13,107	44	13,113	0.80	0.51 – 1.24
Canada (CNBSS2)	38	19,711	39	19,694	0.97	0.62 – 1.52
subtotal (n=2)			fixed		0.88	0.64 – 1.20
			random		0.88	0.64 – 1.20
			Q		p = 0.530	
Kopparberg	59	29,426	44	13,793	0.63	0.43 – 0.93
Östergötland	42	28,722	57	27,311	0.70	0.47 – 1.04
Stockholm	33	25,476	28	12,840	0.59	0.36 – 0.98
Gothenburg b	21	9,903	37	15,708	0.90	0.53 – 1.54
subtotal (n=6)			fixed		0.74	0.62 – 0.89
			random		0.74	0.62 – 0.89
			Q		p = 0.616	
New York	52	16,151	80	16,089	0.65	0.46 - 0.92
total (n=7)			fixed		0.72	0.61 - 0.85
			random		0.72	0.61 - 0.85
			Q		p = 0.674	

Table C2 As table C1, except observation period: thirteen years.

trial	intervention group		control group		RR	95% CI
	Obs	N	Obs	N		
Malmö	79	17,430	92	17,426	0.86	0.64 – 1.16
Canada (CNBSS2)	107	19,711	105	19,694	1.02	0.78 – 1.33
subtotal (n=2)			fixed		0.94	0.77 – 1.15
			random		0.94	0.77 – 1.15
			Q		p = 0.406	
Kopparberg	104	29,007	88	13,551	0.55	0.42 – 0.73
Östergötland	112	28,229	150	26,830	0.71	0.56 – 0.91
Stockholm	42	25,476	33	12,840	0.64	0.41 – 1.01
subtotal (n=5)			fixed		0.75	0.66 – 0.86
			random		0.75	0.60 – 0.93
			Q		p = 0.028	
New York	101	16,505	130	16,505	0.78	0.60 – 1.01
total (n=6)			fixed		0.76	0.67 – 0.85
			random		0.75	0.63 – 0.90
			Q		p = 0.052	

Annex **D**

All-cancer mortality

See table on the following page.

Table D Deaths due to any cancer, including breast cancer (obs), number of women (N) and relative risk (RR), with 95 per cent confidence interval (95% CI) for death from cancer in the intervention group (i.e. those invited to mammographic screening), compared with the control group (i.e. those not invited). All women aged forty to seventy-four at the start of the trial.

trial	intervention group		control group		RR	95% CI
	Obs	N	Obs	N		
Malmö	707	21,088	739	21,195	0.96	0.87 – 1.06
Canada (NCBSS1)	281	25,214	283	25,216	0.99	0.84 – 1.17
Canada (NCBSS2)	464	19,711	403	19,694	1.15	1.01 – 1.31
subtotal (n=3)				fixed	1.02	0.95 – 1.10
				random	1.03	0.92 – 1.15
				Q	p = 0.101	
Kopparberg	666	39,051	319	18,846	1.01	0.88 – 1.15
Östergötland	510	39,034	498	37,936	1.00	0.88 – 1.13
subtotal (n=5)				fixed	1.01	0.96 – 1.07
				random	1.01	0.95 – 1.08
				Q	p = 0.317	
New York	791	30,239	823	30,765	0.98	0.89 – 1.08
total (n=6)				fixed	1.00	0.96 – 1.05
				random	1.00	0.96 – 1.05
				Q	p = 0.403	

Annex **E**

Overall mortality

See tables on the following page.

Table E1 Deaths due to any cause (obs), number of women (N) and relative risk (RR), with 95 per cent confidence interval (95% CI) for overall mortality in the intervention group (i.e. those invited to mammographic screening), compared with the control group (i.e. those not invited to mammographic screening). Women aged fifty (Malmö fifty-five) or older at the start of the trial. Observation period: seven years.

trial	intervention group		control group		RR	95% CI
	Obs	N	Obs	N		
Malmö	253	19,711	250	19,694	1.01	0.85 - 1.20
subtotal (n = 1)			fixed		1.01	0.85 - 1.20
Kopparberg	3,485	29, 007	1,619	13,551	1.01	0.95 - 1.06
Östergötland	3,385	28,229	3,332	26,830	0.97	0.92 - 1.01
Stockholm	1,494	24,836	864	12,957	0.90	0.83 - 0.98
Göteborg b	349	10,112	591	15,997	0.93	0.82 - 1.06
total (n=5)			fixed		0.97	0.94 - 1.00

Table E2 As table E1, except observation period: thirteen years.

trial	intervention group		control group		RR	95% CI
	Obs	N	Obs	N		
Malmö	2,361	17,101	2,423	17,128	0.98	0.93 - 1.03
Canada (NBSS 2)	734	19,711	690	19,694	1.06	0.96 - 1.18
subtotal (n = 2)			fixed		0.99	0.95 - 1.04
			random		1.01	0.93 - 1.09
			Q		p = 0.145	
Kopparberg	5,725	28,918	2,659	13,470	1.00	0.96 - 1.05
Östergötland	4,564	28,657	4,398	27,216	0.99	0.95 - 1.02
total (n = 4)			fixed		0.99	0.97 - 1.02