as far as cancer is concerned because human cancers are unique and not accurately mimicked by animal models. Thus the effects on tumour and on normal tissues cannot be predicted in more than general terms from animal studies. There is, therefore, a need to define appropriate preclinical testing for phase I trials. Information from such clinical studies is a suitable test for many of the new products of recombinant DNA technology because they are designed specifically for actions in man which are not accurately reproduced in animal models.

The Cancer Research Campaign Phase I/II Trials Committee has addressed similar problems in development of cytotoxic drugs and monoclonal antibodies, producing operation manuals of control recommendations for products prepared for investigational administration to patients with cancer in phase I trials. This group, together with the National Institute for Biological Standards and Control and the EORTC Group for Biological Agents has now produced an operation manual for products of recombinant DNA technology which is published in this issue (pp. 1907–1910). This defines suitable preclinical testing which is within the scope of academic departments or capable of being contracted out at a moderate cost.

With the route for phase I studies facilitated in this way, clinical investigators can more easily address the challenging problems of characterisation of the biological effects of these exciting new agents in man.

R. Begent
Department of Clinical Oncology
Royal Free Hospital School of Medicine
Rowland Hill Street
London NW3 2PF
U.K.

Eur J Cancer, Vol. 29A, No. 13, pp. 1804–1805, 1993.
Printed in Great Britain

0959–8049/93 $6.00 + 0.00

Firmer Evidence on the Value of Breast Screening—the Swedish Overview

Jocelyn Chamberlain

MORE HAS been done in Sweden to evaluate breast cancer screening than in any other country. No fewer than five randomised controlled trials have been carried out between 1976 and 1990, in Kopparberg, Ostergotland, Malmo, Stockholm and Gothenburg. The Kopparberg and Ostergotland trials, together known as the Two Counties Study, published their first results in 1985, showing a 31% reduction in breast cancer deaths 7 years after women were first invited for screening [1]. This positive outcome, together with similar findings from the original screening trial conducted 10 years earlier in New York [2], and three retrospective studies also indicating a reduction in risk of breast cancer death in screened women [3–5], led many authorities to conclude that breast cancer screening had been adequately tried and tested and was now ready for widespread implementation [6].

But several months later, consternation was caused by publication of Malmo results which showed only a 4% non-significant mortality reduction, appearing only *after* the scheduled 'end' of the trial, 10 calendar years from its start, equivalent to an average period of follow-up for each woman of 8.8 years [7]. Much emphasis was placed in this report on validation of the cause of death among patients with breast cancer. An independent death review committee found that for 15 out of 193 deaths in patients with breast cancer there was a possibility of biased classification, and with a difference of only three deaths between the invited and control groups such a level of misclassification could have led to biased conclusions about the effectiveness of screening.

Recognising that the Malmo study lacked statistical power because of its small sample size, and that this problem could to a lesser extent also constrain the Stockholm and Gothenburg studies, the Swedish Cancer Society initiated an independent review of combined data from all five trials in order to reach a more definitive answer on the extent to which screening could reduce breast cancer mortality in Swedish women [8].

A team was appointed to conduct the overview headed by a senior oncologist, and comprising epidemiologists and clinicians who had not been involved in any of the trials. Unlike meta-analyses which aggregate 'raw' ungrouped data from a number of trials, or overviews which sum up already published data, this team chose a different approach. Their method was specifically designed to ensure consistency between the five trials (a) in the completeness of ascertainment of the main end-point — death from breast cancer; (b) in its classification by an independent committee; and (c) in the method of analysis. The only data from the original trials submitted to the overview was a register of every woman in the trials, including identification details, date of birth, date of randomisation, and randomisation group — invited or control. These registers were then merged into one and matched against the Swedish national death register (100% complete) and cancer registry (98% complete). From these sources all 1301 women who had been diagnosed with breast cancer after their date of entry to the trial, and who had died of any cause before 1 January 1990, were selected for special study.

Copies of their death certificates, autopsy reports, pathology reports and medical records were obtained; some clinical information about the period leading up to death was available for 99% of cases. The relevant records, blinded as to identity and randomisation group, were then reviewed by an end-point committee consisting of the pathologist, surgeon, oncologist and radiologist from the overview team. The deaths were classified as follows: (i) breast cancer was the underlying cause of death; (ii) active breast cancer was present at death although not necessarily the underlying cause; or (iii) active breast cancer was not present at the time of death. Each member of the end-point committee assessed each case independently and their

Correspondence to J. Chamberlain at the DH Cancer Screening Evaluation Unit, The Institute of Cancer Research, Section of Epidemiology, 15 Cotswold Road, Sutton, Surrey SM2 5NG, U.K.
Received 18 June 1993; accepted 29 June 1993.

classifications were unanimous in 90% of all cases. In the remaining 10% a consensus was reached after discussion.

Two end-points — breast cancer as the cause of death, and breast cancer present at death — were then expressed as a rate per 1000 woman-years of follow-up, and the relative risk of the invited group compared with the control group was calculated for each end-point. The situation was complicated by the fact that in all the trials except Malmo, the control group was also invited for screening between 1982 and 1989, 3 to 5 years after womens' entry in to the trial. In Malmo, control women were not specifically invited but it is known that 25% of them had mammography during the course of the trial. Should the comparison be restricted to deaths of patients whose cancer was diagnosed between date of entry and the date of completion of screening the control group, in order not to dilute the effect in the invited group? Or should deaths in women whose cancer was diagnosed after the controls were screened be included in the comparison, since in the invited group some of these cases will have been influenced by screening, and in any case little effect on mortality would be expected in the first 4–5 years after starting to screen controls? In the event, the overview chose both these approaches, the first being named 'the evaluation model' and the second 'the follow-up model'. Fortunately for interpretation there was very little difference in the relative risk between the two end-points and between the two analysis models.

The results of these new analyses add weight to earlier conclusions about the effectiveness of screening. For the total sample of women aged 40–74, the relative risk of an invited woman dying from breast cancer compared with a control group woman was 0.76 using the evaluation model. With a total number of woman-years of follow-up in excess of 2.5 million, the 95% confidence intervals around the relative risk estimate were narrow (0.66–0.87) and the result highly significant. The relative risks in individual trials ranged from 0.68 in Kopparberg to 0.84 in Ostergotland and Gothengburg, but confidence intervals were wide and the only statistically significant finding was in Kopparberg.

When analysed by age at entry, a statistically significant 29% reduction in mortality was found for woman aged 50–69, but in those aged 40–49 it was a non-significant 13% (RR 0.87, 95% confidence interval 0.63–1.20), and in women aged 70–74 only 6% (RR 0.94, 95% confidence interval 0.60–1.46). An analysis of cumulative breast cancer mortality over time showed that the cumulative mortality in the invited group began to diverge from that in the control group around 5–6 years after trial entry and continued to diverge up to the maximum follow-up of 12 years. In women aged 50–59 the divergence started earlier, at 4–5 years from entry, while in women aged 40–49 it did not appear until 8 years from entry. This latter finding mimics that of the earlier trial in New York [2]. There was no suggestion of an excess mortality in invited women in this younger age group in the early years of the trial, as has been suggested by others.

This overview was a difficult, expensive, and, above all, time-consuming exercise. To retrieve clinical records for 1300 women years after their death was no slight undertaking, and it is not surprising that it took some 5 years for them all to be reviewed separately by five busy clinicians. So what has it achieved? First and foremost it has confirmed, beyond reasonable doubt, that breast screening between the ages of 50 and 69 reduces a woman's chance of dying from breast cancer over the ensuing 12 years by at least 25–30%. This has been done by the accumulation of a very large sample size coupled with a rigorous method for determining the number of end-points, which is, and more importantly is seen to be, independent of those organising the trials. Secondly, it has shown that even a sample size which included 770 000 woman-years was inadequate to give a clear answer on the level of effectiveness of screening starting before age 50, and that if any true benefit from screening 40–49-year-old women exists, its appearance is likely to be delayed for 8 years from the start of screening.

Thirdly, this combined analysis and independent collection of end-point data refute arguments between proponents of the different trials about spurious explanations for their different findings.

But the overview also raises some tantalising questions about the discrepancy in numbers of end-points between its findings and those reported in published papers from the individual trials. This is particularly true in relation to the findings in women aged 40–49, where the overview's point estimate of relative risk (0.87) is lower than that of Malmo (1.29) [7], the combined Two Counties study (0.98) [9] and Stockholm (1.07) [10]. The overview did not give a breakdown of its findings by age group within centre, so one can only guess at possible explanations. These include differing periods of follow-up, differing classifications of causes of death, and differing classifications of age group. The overview used date of birth to calculate exact age at entry whereas the trials used year of birth. And in Malmo there was no separation of 45 to 54 year grouping into those aged 45 to 49 at entry and those aged over 50. Nor can the overview contribute any combined data about breast cancer deaths in the invited group in relation to screening history, because data on screening episodes were not transferred to the overview team.

Whether these points can be addressed in the future remains to be seen, and are of low priority in comparison with the main task which has already been achieved. For populations with high breast cancer mortality, and with plentiful resources, provision of mammographic screening for women aged 50–69 is confirmed as a sensible public health policy.

1. Tabar L, Gad A, Holmberg LH, *et al.* Reduction in mortality from breast cancer after mass screening with mammography. *Lancet* 1985, i, 829–832.
2. Shapiro S. Evidence on screening for breast cancer from a randomized trial. *Cancer* 1977, **39**, 2772–2782.
3. Collette HJA, Day NE, Rombach JJ, de Waard F. Evaluation of screening for breast cancer in a non-randomised study (the DOM Project) by means of a case-control study. *Lancet* 1984, i, 1224–1225.
4. Verbeek ALM, Hendricks JHCL, Holland R, Mravunac M, Sturmans F, Day NE. Reduction in breast cancer mortality through mass screening with modern mammography. (First results of the Nijmegen Project 1975-81.) *Lancet* 1984, i, 1222–1224.
5. Palli D, Rosselli del Turco M, Buiatti E, *et al.* A case-control study of the efficacy of a non-randomised breast cancer screening program in Florence (Italy). *Int J Cancer* 1986, **38**, 501–504.
6. Day NE, Baines CJ, Chamberlain J, Hakama M, Miller AB, Prorok P. UICC Project on Screening for Cancer. *Int J Cancer* 1986, **38**, 303–308.
7. Andersson I, Aspegren K, Janzon L, *et al.* Mammmographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *Br Med J* 1988, **297**, 943–994.
8. Nystrom L, Rutqvist LE, Wall S, *et al.* Breast cancer screening with mammography. *Lancet* 1993, **341**, 973–978.
9. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiologic Clinics of N America* 1992, **30**, 187–210.
10. Rutqvist LE, Miller AB, Andersson I, *et al.* Reduced breast cancer mortality with mammography screening — an assessment of currently available data. *Int J Cancer* 1990, (Supplement 5), 76–84.