The Swedish Two-County Trial of mammographic screening: cluster randomisation and end point evaluation

S. W. Duffy¹*, L. Tabar², B. Vitak³, M. F. Yen¹, J. Warwick¹, R. A. Smith⁴ & H. H. Chen⁵

¹Queen Mary University of London, London, UK; ²Falun Central Hospital, Falun, Sweden; ³University of Linköping, Linköping, Sweden; ⁴American Cancer Society, Atlanta, GA, USA; ⁵Institute of Preventive Medicine, National Taiwan University, Taipei, Taiwan

Received 19 July 2002; accepted 3 April 2003

Introduction

The randomised trials of breast screening have demonstrated a substantial and significant reduction in breast cancer mortality in association with invitation to mammographic screening [1, 2]. Recently, the evidence has been called into question in a meta-analysis by Olsen and Gøtzsche [3, 4], who maintained that poor quality of the trials and lack of validity of breast cancer mortality as an end point cast doubt on the evidence for a benefit. In terms of the Swedish Two-County Trial, Olsen and Gøtzsche asserted, on the basis of a slight age imbalance between study and control groups, that the cluster randomisation was inadequate and that the cause of death classification was invalid, on the basis of an alleged excess of deaths from other causes among the breast cancer cases diagnosed in the active study population (ASP) and PSP.

Although these and other alleged shortcomings in the Two-County Trial have been shown to be ill-founded [5, 6], we therefore address the issues of potential biases resulting from the cluster randomisation and the effect of invitation to screening on other causes of death among the breast cancer cases. In approaching the latter problem, we take account of the additional time for study group cases to die of other causes due to lead time.

Patients and methods

The Swedish Two-County Trial is a cluster-randomised trial of mammographic screening for breast cancer, conducted in Östergötland (E) and Kopparberg (W) counties in Sweden. Excluding those with breast cancer before randomisation, 77,080 women aged 40–74 years were randomised in geographical clusters to invitation to screening (active study population, ASP) and 55,985 to no invitation (passive study population, PSP), in 45 geographical clusters. After ~7 years, the PSP was invited to screening and the trial closed. We analysed data using hierarchical statistical models to take account of cluster randomisation, and performed a conservative analysis assuming a systematic difference between ASP and PSP in baseline breast cancer mortality in one of the counties. We also analysed deaths from causes other than breast cancer and from all causes among breast cancer cases diagnosed in the ASP and PSP.

Conclusions: Breast cancer mortality is a valid end point and mammographic screening does indeed reduce mortality from breast cancer. The criticisms of the Swedish Two-County Trial are unfounded.

Key words: breast screening, mammography, randomised trial

© 2003 European Society for Medical Oncology
Results

Up to the end of the first screen of the PSP there were 1426 cancers diagnosed in the ASP and 1042 in the PSP. Table 1 shows the age distribution of women in the ASP and PSP, and the numbers of breast cancer deaths to 1998 by age. Although the age distributions are very similar, and the difference would tend to bias the result against screening, the small difference is significant, due to the large numbers. This may result from the cluster randomisation, so analysis that takes account of the cluster randomisation is advisable. Adjusting for age in the analysis is also necessary. The age-adjusted reduction in breast cancer mortality associated with screening was 31% [relative risk (RR) = 0.69, 95% CI 0.58–0.80, \( P < 0.001 \)].

Table 2 shows the results from the hierarchical models 1–4, age-adjusted. The results of all four models are very similar, all significant and showing an \( \approx 30\% \) mortality reduction, indicating that the cluster randomisation was efficient. To take account of possible systematic differences in baseline mortality between clusters, we considered breast cancer deaths within clusters in the 10 years before the trial. There was no difference in the rates between ASP and PSP clusters in E county [2]. Due to changes in the tax districts, we could not obtain pre-trial mortality by cluster within three of the strata representing large municipalities in W county, but we did have the data for the 12 small municipality clusters within the remaining four strata. These showed a non-significant 15% lower rate of breast cancer mortality in the ASP clusters in the 10 years before the trial. Although this is compatible with pure chance, we assumed conservatively that it was real and that it also applied to the three remaining strata. Adjusting for the pre-trial differences in mortality, there remained a significant reduction in breast cancer mortality in the ASP, albeit slightly attenuated at 27% (RR = 0.73, 95% CI 0.63–0.85, \( P < 0.001 \)).

Adjusting for age and taking account of the longer follow-up time in the ASP, there was no significant difference between deaths from other causes among the breast cancer cases in the ASP and PSP (RR = 1.12, 95% CI 0.96–1.31, \( P = 0.14 \)). There was a significant 13% reduction in deaths from all causes in breast cancer cases in the ASP (RR = 0.87, 95% CI 0.78–0.99, \( P = 0.02 \)).

Discussion

The above results show no evidence of bias, either from the cluster randomisation or from misclassification of cause of death. The most conservative analysis based on an assumed systematic difference in baseline mortality in W county still gave a significant 27% reduction in breast cancer mortality in the ASP. Ignoring the cause of death classification, there was a significant 13% reduction in all-cause mortality among breast cancer cases in the ASP. This suggests that the classification of cause of death is reliable, and that death from breast cancer is the appropriate end point.
These results are not surprising. We have dealt with the issues of cluster randomisation and cause of death in our previous publications [8, 9, 11–13]. In addition, the Swedish overview has published an excess mortality analysis that did not rely on classification of cause of death and which came to the same conclusions as the breast cancer mortality analysis [14].

Concerns about the screening of the PSP or its timing should be allayed by two factors. First, the published result in 1985 (of breast cancer mortality to the end of 1984) is essentially the pre-PSP screening result. Although the planned first screen of the PSP had started by the end of 1984, it was only 5% complete and had no bearing on mortality at the time. Also, while 13% of the control group had experienced mammography by this time, <2% had received screening mammography (i.e. true contamination). Secondly, the Swedish overview [15] analysed the results by the follow-up model including all breast cancers diagnosed up to the end of 1989, and obtained a similar result to ours for the Two-County Trial.

We therefore conclude that the major criticisms of the Two-county Trial are unfounded. The trial is not perfect, nor is any other study. Breast cancer mortality is a valid end point and mammography screening does reduce deaths from breast cancer.

Acknowledgements

We thank the women who took part in the Two-County Trial and the large number of clinical and support staff who contributed their time and talents to it.

References