Mammographic screening: case–control studies

S. D. Walter*

Clinical Epidemiology and Biostatistics, McMaster University, Health Sciences Centre, Hamilton, Ontario, Canada

Received 4 July; accepted 3 April 2003

Background: The case–control design can be used to evaluate the benefit of cancer screening programmes.

Materials and methods: This paper outlines the main methodological features of the case–control design in this context, and indicates some potential biases. It also reviews the existing case–control literature on mammographic screening.

Results: Case–control studies consistently indicate a reduction of −50% in breast cancer mortality associated with mammography. This result indicates greater benefit than shown in randomised trials; however, one should recognise that trials indicate effectiveness whereas case–control studies indicate efficacy. The two types of evidence are broadly compatible when one allows for screening non-compliance and contamination in the randomised trials.

Conclusions: The case–control evidence supports and is consistent with the findings of randomised trials of mammography. Effectiveness estimates from trials indicate the benefit of screening to the population as a whole, and are pertinent to the public policy debate as to the value of offering screening. In contrast, case–control studies indicate benefit to actual screening participants. As such, case–control estimates of efficacy are appropriate for individual decision-making by women about their use of mammography when it is potentially available to them.

Key words: case–control studies, effectiveness, efficacy, mammography, trials

Introduction

The case–control design can be used to evaluate cancer screening, and has certain advantages over prospective designs such as cohort studies or randomised trials. Specifically, case–control studies can often be completed rapidly, at low cost and with high statistical efficiency. However, the case–control design also has a number of disadvantages. The methodology for the analysis is more complex and perhaps less familiar to consumers, and certain biases may affect the results [1–3].

The goal of mammographic screening is to reduce breast cancer mortality. In a randomised trial, evaluation is achieved by comparing an experimental group of women offered screening with control women not offered screening. In contrast, in the case–control design the evaluation is by a retrospective comparison of the screening history in cases and controls. For a valid design, an investigator must consider the appropriate definition of the cases, of the controls and of the screening history.

Materials and methods

Case definition

The cases are specified according to the event that the screening is intended to prevent, which for mammographic screening is death from breast cancer. Hence the cases should consist of deaths from breast cancer during a defined time period.

Control definition

The controls should represent the screening history in the population in which the cases arise. They should be at risk of breast cancer death, and optionally might be matched for factors such as age. A control should be alive when her matched case dies, but she should not have been diagnosed with breast cancer before the case, because otherwise she would no longer be eligible for screening.

Screening history

It is necessary to count screening examinations that can potentially lead to early diagnosis, more effective early therapeutic intervention and reduced mortality. Specifically, these screening tests will be those that occur during the preclinical period of cancer development, and when the disease is screen-detectable.

If the screening test is sensitive enough to detect disease, and if effective treatment is initiated, then benefit from screening will be evident by screening being observed more commonly in controls. The relative frequency of screening is assessed through the case–control odds ratio, which is <1 if screening is beneficial.

It is necessary to distinguish true screens from diagnostic tests; in other words, the only screens to be counted should be those carried out at times when the patient was asymptomatic. The screening history should include tests with both positive and negative results. Tests carried out after the patient’s diagnosis should be excluded, because they cannot be true screens. For reasons of comparability, screens carried out on the control after the date of the case diagnosis should also be excluded. Finally, screening tests that identify cases should be included.
Potential biases affecting case–control studies

Investigators sometimes limit their cases to advanced-stage disease. However, this will exclude some cases of early-stage disease who are more likely to be screen-detected. This reduces the screening rate among cases and creates a pro-screening bias.

A related problem occurs if early-stage cases are used as controls. This strategy is sometimes used in clinic-based studies, where disease-free controls are not available. This approach may lead to higher screening rates being observed among the controls, for reasons of stage shift, length bias and lead-time bias. Increasing the rate of screening in the controls again leads to a pro-screening bias.

As mentioned earlier, both positive and negative screening test results should be used in the analysis. The goal of the evaluation is to consider the use of the test, and the result cannot be predicted before the test is carried out. Positive tests are more likely to occur in cases, so restriction of the analysis to only positive test results would produce a bias against screening.

Inclusion in the analysis of the index test identifying screen-detected cases does lead to an anti-screening bias, the magnitude of which depends on the screening interval, the duration of the preclinical disease period, and the extent of screening coverage in the population. The reason is that cases can be screened until they become screen-detected or diagnosed clinically as an interval case. However, the control’s exposure can only be counted until the date of case diagnosis, which therefore means less exposure opportunity. If the index screens are excluded, there is a bias in favour of screening, so the conservative approach of including the index screen seems preferable. Numerical investigation suggests that the magnitude of the resultant bias is modest [4, 5].

Case–control evidence on mammography

A MEDLINE search revealed one meta-analysis by Demissie et al. [6], involving five case–control studies [7–14], and two more recent case–control studies [15, 16]. There were also a larger number of studies involving comparisons such as between early- and late-stage cancers, or between screen-detected and clinically detected cases with respect to their clinical features. These latter studies do not consider breast cancer mortality, so they cannot be used directly to provide an estimate of screening benefit. Therefore, we will focus here on the mortality-based studies, and by doing so we avoid most of the biases mentioned above.

Table 1 shows summary statistics for the five case–control studies used in the meta-analysis and Table 2 shows the principal results. The overall result shows a benefit of screening with an odds ratio of 0.44 (95% CI 0.38–0.50), and there was no significant heterogeneity between studies. The odds ratios for women aged 40–49 years at entry indicate somewhat less benefit, but the confidence intervals are wide and only two studies considered this age subgroup.

Demissie et al. [6] compared the results from the five case–control studies with nine randomised trials. The overall odds ratio from the trials was 0.76 (95% CI 0.69–0.83), indicating less protection than in the case–control studies. The overall ratio of the trial effect to the case–control effect was 1.74 (95% CI 1.48–2.04). There are several possible explanations for the larger effect in the case–control studies, including their longer follow-up and contamination between groups in the trials (e.g., controls might actually be screened). However, the most likely explanation has to do with the effect measure implied by each design.

Randomised trials evaluate the effect of offering screening to women, compared with controls not offered screening. An intention-to-treat analysis is employed, indicating benefit to the community. The magnitude of that benefit can then be used in debating the public policy of whether screening should be offered. Because this type of analysis allows for possible refusals by women offered screening, and for contamination of the control group, the screening benefit is measured through its effectiveness.

In contrast, the benefit assessed in case–control studies is one of efficacy. Women in the case–control study are classified as to whether they are actually screened or not, and hence measurement is of the benefit of screening among women who actually comply to be screened. Therefore, the efficacy measure is one that is appropriate for individual decision-making.

By making some assumptions about the compliance rate in women offered screening and about the rate of contamination in controls, a ‘conversion’ from effectiveness to efficacy is attempted. Demissie et al. [6] assumed a median compliance rate of 65% and contamination of 20%. When applied to their trials odds ratio of 0.76, they obtained a ‘corrected’ odds ratio of 0.52, which is reasonably close to the value 0.44 observed in the case–control studies. A similar analysis was carried out recently by Duffy and Cuzick [17].

The two more recent case–control studies showed similar results. In the first, a 17-year analysis was carried out for participants in the DOM study [16]; yielding an odds ratio of 0.54 (95% CI 0.37–0.79). The second involved an analysis of older women (≥65 years) in Nijmegen [15], and gave an odds ratio of 0.56 (95% CI 0.28–1.13). Both of these estimates are compatible with the earlier results.

Discussion

The methodology of case–control evaluation of screening is complex, and there are several potential pitfalls. However, by using breast cancer deaths as the cases and appropriate population controls (as in the studies referred to above), many of the possible biases are avoided. An additional concern in using case–control evidence is that the comparison groups are not randomly assigned. It is necessary to consider selection biases that might convey a different risk of breast cancer mortality among screenees as compared to non-screenes. Several of the studies in the meta-analysis [6] investigated this but found no evidence of selection bias.

One topic that does not appear to have received much attention is the need to distinguish screening from diagnostic mammography. In the evaluation of cervical Pap smear screening, for
instance, the application of different rules defining which tests are counted as screens can affect the estimated benefit. Additionally, there may be differences in how tests are classified according to the source of information (clinical records or personal interviews) on symptoms [18, 19].

This issue appears not to have been explored for breast cancer. If a mammography that is part of a diagnostic process is inappropriately counted as screening, one might expect a bias against screening, because the rate of screening among the cases would be inflated. Further work is needed in this area, as well as to explore more fully the pattern of screening at particular times during a woman’s history, and with particular frequencies.

Despite these concerns, the case–control evidence on mammography appears to be broadly consistent with the trials. Among women who comply with screening, the efficacy measure of benefit is a reduction in breast cancer mortality of ~50%.

### References


### Table 2. Case–control studies of mammography: results (odds ratio and 95% confidence interval)

<table>
<thead>
<tr>
<th>Study</th>
<th>All women</th>
<th>Women aged 40–49 years at entry</th>
<th>Women aged 50–74 years at entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOM</td>
<td>0.30 (0.13–0.71)</td>
<td>–</td>
<td>0.30 (0.13–0.71)</td>
</tr>
<tr>
<td>Nijmegen</td>
<td>0.51 (0.26–0.99)</td>
<td>1.23 (0.31–4.81)</td>
<td>0.26 (0.10–0.67)</td>
</tr>
<tr>
<td>Florence</td>
<td>0.57 (0.35–0.92)</td>
<td>0.83 (0.37–1.85)</td>
<td>0.49 (0.26–0.89)</td>
</tr>
<tr>
<td>UK</td>
<td>0.51 (0.27–0.98)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Malmö</td>
<td>0.42 (0.22–0.78)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Overall</td>
<td>0.44 (0.38–0.50)</td>
<td>0.88 (0.47–1.63)</td>
<td>0.32 (0.28–0.38)</td>
</tr>
</tbody>
</table>