Letters to the Editor

Misleading quotations and other errors persist in rejoinder on breast cancer screening

From PETER C GÔTZSCHE

Sirs—Although I demonstrated that the review by Freedman, Petitti and Robins (FPR) of breast cancer screening1 contains many errors,2 they persist.3 In their review,1 FPR noted that 434 more women with breast cancer prior to randomization were excluded from the study group than from the control group in the Health Insurance Plan (HIP) trial, quoting the principal investigator of this trial. However, in their rejoinder,3 they claim this number originated with me. I wrote that many more women with breast cancer prior to randomization were excluded from the study group than from the control group2 and FPR contradict themselves when they now say my statement is incorrect.3 FPR also say that I withdrew a previous ‘near-retraction’. There is no ‘near-retraction’ in the reference FPR citation4 where I merely restated my concern that retrospective exclusion of women after 18 years of follow-up may not be reliable.

FPR disregard the large discrepancy in the Two-County study involving a benefit of 24% reported by the trialists versus only 10% reported in the Swedish overview.1 Further, although they consulted extensively with the Two-County trialists, they avoid addressing the many discrepancies in numbers of women and deaths reported for this study.1 FPR also turn a blind eye to overdiagnosis,1 though the recent IARC/WHO report on screening acknowledges this is ‘an obvious source of harm’.5 The report notes that there was about a 50% increase in breast cancer during 5 years after introduction of screening in Finland and UK and that the rise may be persisting,5 in agreement with our findings and those of others.2

FPR avoid discussing length bias3 and do not acknowledge their error in comparing total mortality among breast cancer cases in the study group versus the control group and then concluding that a significant difference is evidence that screening is effective. It is well-known that case-survival is a highly misleading outcome,6 particularly in screening trials.2,7 Finally, FPR reiterate that it is appropriate to exclude deaths in women who have been invited to screening but refused to get screened. This is not how trials should be analysed if we wish to avoid bias (www.consort-statement.org).

References


In search of the best available version of the truth

From CORNELIA J BAINES

Sirs—In their qualitative assessment of screening trials Freedman et al.1 demonstrate that impressive author affiliations and the peer review process do not guarantee accuracy. On page 50, the reader will find the following persuasive statement about the Canadian National Breast Screening Study (CNBSS). ‘Centre radiologists only agreed with the reference radiologist 30–50% of the time.’ What a dreadful study that must have been! Quite wrong.

The numbers they cite are clearly reported as kappa statistics, which indicate how much of the agreement observed was agreement beyond that which might occur by chance.2 In fact, Table 2 (which they disregarded) from our publication reveals...
very clearly that there was agreement between centre radiologists and the reference radiologist 85.6% of the time with respect to cancer cases and 75.8% of the time for mammograms from women who did not have cancer. Do these authors truly not understand the difference between inter-observer agreement and kappa statistics?

They go on to condemn the Canadian study for false negative mammograms. ‘Observer error and technical problems led to delayed detection in 22–35% of cancers.’ Our paper reports that the overall false negative rate was 24%: 22% for screen cancers and 35% for interval cancers; by not differentiating between screen and interval cancers, a false impression is conveyed. Contrast those rates with words from the radiologist D Kopans who wrote of a clinical study, ‘this review confirms another well known phenomenon, namely the failure of even expert observers to perceive all abnormalities.’ He continues that a radiological review discovered that 54% of cancer cases had been read as negative. Although more than twice the false negative rate in the CNBSS, somehow Dr Kopans found 54% acceptable. It was ‘inflated’ because the review was not blind.

Page 51 begins with an attribution to me which is uninterpretable (‘Baines notes that a comparison of advanced cancers detected by MA + PE in treatment to those detected by PE in control (19 to 5) is biased.’) and cannot be identified on page 329 as is claimed. As for allegations about flawed mammography and subverted randomization, these are stale, unwarranted and have been responded to previously.

Not only did the authors present false data but they also based their evaluation on CNBSS results published 12 years ago rather than results published in 2000 and 2002, Table II. This must compel any intelligent reader to question the validity of their conclusions. Unfortunately, it is not the first time, nor will it be the last, that false data about the CNBSS have been published.

The authors acknowledge spending much time in consultation with Dr L Tabar of the Two-County trial, presumably to help them understand Swedish data. It is unfortunate that they did not make similar efforts to understand Canadian data.

References


Authors’ response

From DAVID A FREEDMAN, DIANA B PETITTI and JAMES M ROBINS

We reply to a few of the points raised in the letters, starting with Baines. Kappa is a widely-used measure of inter-observer agreement, which corrects raw agreement rates for coincidental matches. For example, if a coin lands heads with probability 50%, two flips of the coin will match 50% of the time, just by chance. The raw agreement rate is 50%. Kappa is 0, because there is no agreement above the chance level.

Baines et al. used Kappa to measure agreement rates between the ‘reference radiologist’ and radiologists at the centres where screening was actually done. We followed suit. Baines now seems to deny that raw agreement rates are inflated, or that Kappa is a measure of agreement.

The radiology in the Canadian National Breast Screening Study (CNBSS) has been widely criticized. Poor radiology would explain the lack of effect in CNBSS. Extensive disagreements between the reference radiologist and centre radiologists support that explanation. The critics have additional data, and the CNBSS investigators defend the study. On balance, we think CNBSS had problems.

Baines reiterates a point made by Miller, that we ignored later CNBSS publications. This is wrong (ref. 3, p. 51; ref. 6, p. 70). We ask Baines to consider our discussion of such publications. There is an excess of advanced cancer detected by physical examination at baseline, among women age 40–49 assigned to screening in CNBSS. Miller and Gotzsche explain the excess on the theory that these were small cancers, which were hard to find when women were referred ondiagnosis work-up and diagnostics were more thorough in the screening arm. Given recently published CNBSS data on tumour size, this explanation is not tenable. The tumours were not small.

Gotzsche says ‘434 more women with breast cancer prior to randomization were excluded from the study group than from the control group in the Health Insurance Plan (HIP) trial...’ He says this is our discovery. He forgets that four years ago, he wrote (ref. 9, pp. 129–30):

Women were excluded if breast cancer had been diagnosed before entry to the trial, and this status was more completely