Screening effectiveness

Outcome events in studies of diagnostic or screening tests

N S Weiss

How should they be defined?

Judgements about the effectiveness of most diagnostic or screening tests in leading to improved health outcomes typically incorporate data from several different types of research.[1] For example, the magnitude of the benefit of screening diabetic patients for retinopathy can be estimated from combining the results of separate studies that estimate: (*a*) the frequency of positive findings on retinal photography or on ophthalmoscopy; (*b*) the accuracy of these methods in identifying retinopathy and in predicting visual loss; and (*c*) the efficacy of photocoagulation in retarding loss of vision.

In some circumstances, however, the only way to judge the potential ability of a test to lead to improved outcome is by directly comparing, in a single study, the experience of people who do and do not receive the test with regard to the occurrence of the outcome that testing is seeking to avert. This is commonly the case when evaluating the effectiveness of cancer screening, as generally it is not possible to conduct a valid study that corresponds to the efficacy of photocoagulation in retarding loss of vision—that is an assessment of the efficacy of treatment for screen detected cancer. Examples of studies that monitor outcomes in people who do and do not receive a given test, termed "one step studies"[1] to distinguish them from the multistep approach that has been used to investigate the effectiveness of such tests as retinal photography or ophthalmoscopy, include randomised controlled trials of mammography, clinical breast examination, or breast self examination relative to mortality from breast cancer.[2–5] These studies will find a decreased rate of an adverse outcome among screened people only if the test is accurate in identifying a disease precursor or the presence of preclinical disease; and the treatment given after a positive test is more effective than that which would be given later for clinically evident disease.

For most one step studies of screening effectiveness, it is straightforward to define the outcome that screening is intending to avert. In a study of the effectiveness of mammography—for example, that outcome would be death that occurred as a result of breast cancer or the attempts to detect it. (There may be ambiguities in arriving at an operational definition of this, and further ambiguities when trying to decide for individual women whether the operational definition has been met.[6]) However, some studies of screening effectiveness have designated outcome events in a way that makes it difficult or impossible to identify a benefit from testing, should one truly exist. The purpose of this commentary is to delineate the limitations of the particular approaches used in these investigations, and to examine alternative means by which their objectives could be achieved.

EXAMPLE A

The effectiveness of testing is judged by comparing outcomes in all tested and non-tested members of the study population, whether or not they have the condition(s) that testing seeks to uncover.

In a randomised controlled trial in patients with low back pain, Kendrick *et al*[7] assigned half to receive radiography of the lumbar spine and half to receive usual care from their physicians. Nine months later, functional status of the nearly 200 participants enrolled in each arm of the trial was assessed. The group assigned to receive radiography fared somewhat worse for this outcome than the control group, but the difference was modest and statistically compatible with there being no true difference.

Apart from any psychological impact it might have, radiography of the lumbar spine has the potential to influence functional status by virtue of its ability to identify treatable conditions—such as a tumour or an abscess. It has been shown[8 9] that radiography in patients with low back pain identifies a relatively low prevalence of abnormalities of this type, perhaps 5% or less. Therefore, a comparison of an outcome such as functional status in the whole of an intervention group and a control group, in which at least 95% of patients do not have a condition that could have been affected by the intervention measure, will be virtually guaranteed of finding little or no difference no matter how much benefit is achieved in the test positive people.

For a randomised trial of the efficacy of radiography of the lumbar spine to have a reasonable chance of documenting that at least some patients benefited as a result of undergoing this investigation, it would probably have to be one or two orders of magnitude larger than the one conducted by Kendrick *et al*. The number of untoward outcomes occurring in patients with a tumour or abscess in such a study would need to be sufficiently large to enable the reliable detection of a plausible difference in the rate of these outcomes between tested and untested patients. In the absence of a study of this size we would be obliged to adopt a multistep approach, estimating as best we can from several types of research: (*a*) the frequency of tumours, abscesses, etc, on *x* ray films of the lumbar spine in patients with low back pain; (*b*) the accuracy of these *x* ray films in detecting such abnormalities and in predicting adverse outcomes resulting from them; and (*c*) the ability of the treatments administered after radiological detection of tumours, abscesses, etc, to improve upon the natural history of these conditions.

EXAMPLE B

Although assessment of the effectiveness of testing focuses on people with the condition that the test seeks to identify, it fails to restrict its attention to just those outcomes that are (or are likely to be) the result of that condition.

Concato *et al*[10] designed a case-control study to estimate the degree to which screening by means of prostate specific antigen (PSA) or digital rectal examination (DRE) can reduce mortality from prostate cancer. Among men diagnosed with prostate cancer during 1991–5 who had received outpatient care at any of 10 Veteran's Administration Medical Centers during 1989–90, cases for the study were those who died of any cause before 2000. Controls were selected from men who received outpatient care for any reason during 1989–90 at these same centers. When the study is completed, the cases and controls are to be compared for the proportion who had been screened by one or both methods before any clinical suspicion of prostate cancer.

A limitation of the investigator's approach derives from the fact that only about a quarter of deaths in men diagnosed with prostate cancer during life are a direct or indirect consequence of this disease. If screening influences the probability of dying from prostate cancer but not from other causes, then

Abbreviations: PSA, prostate specific antigen; DRE, digital rectal examination

Table 1 PSA screening in controls and in men who died as a result of having prostate cancer

PSA Screening	Cases	Controls	OR	95% CI
Yes	30	200	0.43	0.26 to 0.70
No	70	200		
Total	100	400		

Table 2 PSA screening in controls and in men with prostate cancer who died of any cause

PSA Screening	Cases	Controls	OR	95% CI
Yes	30+150=180	200	0.82	0.61 to 1.09
No	70+150=220	200		
Total	400	400		

an analysis that includes as cases men who died of any cause will produce a falsely low estimate of the relative benefit from screening on mortality from prostate cancer.

The possible size of the underestimation is illustrated. Assume a population in which half the men in a given age range have received a PSA screening test during a period corresponding to the presumed duration of the preclinical phase of the disease. If 30% of men who died as a result of having prostate cancer had been screened during the corresponding period, the results shown in table 1 would be obtained in a study of 100 such fatal cases and 400 controls. The odds ratio (OR) of 0.43 suggests that there was a 57% decrease in prostate cancer mortality in screened men. Table 2 incorporates the other three quarters of the deaths among men with prostate cancer into the calculation of the OR associated with PSA screening. Assuming that PSA screening had no impact on any cause of death other than prostate cancer, the proportion of the additional 300 men who had been screened would be identical to that of controls—that is, 0.50—and this analysis would suggest an overall reduction in mortality of only 18%.

The investigators in this study do plan a "secondary analysis" restricted to men who died of prostate cancer and controls. However, they defend their primary analysis (that includes all deaths in men with prostate cancer) on the grounds that "...it is least prone to error and bias, given the difficulties of attributing cause of death".[9] As shown by the comparison of the ORs in tables 1 and 2, their primary analysis will not provide an unbiased estimate of the impact of PSA or DRE screening on the rate of those causes of death that early detection had the potential to avert. If it were thought that routine identification of deaths due to prostate cancer from the death certificate statement of cause of death is insufficiently accurate, it would be possible to conduct a review of medical records, blinded to screening status, of all men with prostate cancer who died to better identify those who died of this disease.[11]

CONCLUSION

Studies that seek to measure the combined impact of early detection and early treatment of disease face several challenges that can limit their ability to identify a benefit when one is truly present. To avoid unnecessarily enlarging of this number, the analysis of these studies needs to: (*a*) focus on people who have the potential to benefit by early detection; and (*b*) in such people, focus on those outcomes that early detection has the potential to influence.

ACKNOWLEDGEMENT

I am grateful to Drs Peter Cummings, Paul Doria-Rose, and Miriam M Treggiari for their suggestions on an earlier version of this manuscript.

J Med Screen 2002;**9**:52–53

Author's affiliations

N S Weiss, University of Washington, Box 357236, Seattle, WA 98195, USA, and the Fred Hutchinson Cancer Research Center, Box 19024, Seattle, WA 98109, USA

Correspondence to: Professor N S Weiss, Box 357236 Seattle, WA 98195, USA; nweiss@u.Washington.edu

REFERENCES

1 **Weiss NS**. *Clinical epidemiology: the study of the outcome of illness, 2nd ed.* Oxford: Oxford University Press 1996.
2 **Shapiro S**, Venet W, Strax P, *et al*. Ten to 14 year effect of screening on breast cancer mortality. *J Natl Cancer Inst* 1982;**69**:349–55.
3 **Miller AB**, To T, Baines CJ, *et al*. 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**:1490–99.
4 **Thomas DB**, Gao DL, Self SG, *et al*. Randomized trial of breast self-examination in Shanghai: methodology and preliminary results. *J Natl Cancer Inst* 1997;**89**:355–65.
5 **Newcomb PA**, Weiss NS, Storer BE, *et al*. Breast self-examination in relation to the occurrence of advanced breast cancer. *J Natl Cancer Inst* 1991;**83**:260–55.
6 **Black WC**, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst* 2002;**94**:167–73.
7 **Kendrick D**, Fielding K, Bentley E, *et al*. Radiography of the lumbar spine in primary care patients with low back pain: randomised controlled trial. *BMJ* 2001;**322**:400–05.
8 **Rockey PH**, Tompkins RK, Wood RW, *et al*. The usefulness of *x* ray examinations in the evaluation of patients with back pain. *Journal of Family Medical Practice* 1978;**7**:455–65.
9 **Kaplan DM**, Knapp M, Romm FJ, *et al*. Low back pain and *x* ray films of the lumbar spine: a prospective study in primary care. *South Med J* 1986;**79**:811–14.
10 **Concato J**, Peduzzi P, Kamina A, *et al*. A nested case-control study of the effectiveness of screening for prostate cancer: research design. *J Clin Epidemiol* 2001;**54**:558–64.
11 **Richert-Boe KE**, Humphrey LL, Glass AG, *et al*. Screening digital rectal examination and prostate cancer mortality: a case-control study. *J Med Screen* 1998;**5**:99–103.