

Data dredging, bias, or confounding

They can all get you into the BMJ and the Friday papers

On 4 October 2002, women who were moderate drinkers received good news: their risk of breast cancer was not raised, according to a report in the *Lancet* that was widely covered by the British media.¹ The bad news was that smoking at an early age was now implicated as a risk factor for breast cancer. However, after they had enjoyed guilt-free drinks (without cigarettes) for only a few days, on 13 November the message was reversed: alcohol did increase the risk of breast cancer after all, but smoking was declared innocent.² The press release proclaimed "Alcohol, tobacco and breast cancer: the definitive answer." A reader was driven to complain in the letters page of the *Guardian* (14 November 2002): "So let me get this right—alcohol's no good anymore, and if you smoked within five years of getting your periods, that's bad news too. Oh no, that was a couple of weeks ago; smoking's okay now ... Do things stop being bad for us if we just forget about them for a bit, do you think?"

This is a familiar story—so much so that in Bristol we set our medical students the exercise of examining the "health scare of the week" that appears each Friday, generally from a study reported in the *BMJ* or *Lancet*.³

Observational studies propose, RCTs dispose

The widespread perception that epidemiological studies generate conflicting and often meaningless findings⁴ has received support from recent randomised controlled trials, which have failed to confirm even apparently robust findings from observational epidemiological studies. The most topical of these relates to hormone replacement therapy. In 1991 a meta-analysis of epidemiological results relating the use of hormone replacement therapy to the risk of coronary heart disease concluded that it halved the risk, and that the evidence was statistically robust (relative risk 0.50; 95% confidence interval 0.43 to 0.56) and that "overall, the bulk of the evidence strongly supports a protective effect of estrogens that is unlikely to be explained by confounding factors."⁵ Results from randomised controlled trials were, however, very disappointing, with the first large scale trial showing no benefit, confirmed in two subsequent trials, resulting in a pooled odds ratio of 1.11 (0.96 to 1.30).⁶ The apparent cardioprotective effects of hormone replacement therapy that had been found in the observational epidemiology studies were overturned. Again, women were left wondering what they should do.

A similar scenario had previously been played out for the antioxidant vitamin β carotene. Promising epidemiological and laboratory findings led to a paper published in 1981 in *Nature* entitled "Can dietary beta-carotene materially reduce human cancer rates?"⁷ Cancers related to smoking seemed particularly tractable, and by 1990 the answer for lung cancer was a clear yes: Walter Willett, reviewing the observational epidemiological data, concluded that "Available data thus strongly support the hypothesis that dietary carotenoids reduce the risk of lung cancer."⁸ Four years later a large scale randomised controlled trial showed an 18% increase (3% to 36%) in lung cancer in those

taking β carotene.⁷ Vitamin E and coronary heart disease provided another example of observational studies and randomised controlled trials failing to reach the same conclusion.⁹

"Eating fruit halves the risk of an early death" the *Independent* claimed¹⁰ in an excited response to a study showing a strong inverse association between blood vitamin C levels and mortality due to coronary heart disease.⁸ A subsequent randomised controlled trial of a vitamin supplement that raised blood vitamin C levels by 15.7 $\mu\text{mol/l}$ found five year mortality due to coronary heart disease unchanged (relative risk 1.06; 0.95 to 1.16),⁹ whereas the equivalent observational findings for this increase in blood vitamin C were coronary heart disease relative risks of 0.63 (0.49 to 0.84) in women and 0.72 (0.61 to 0.86) in men (see fig A on bmj.com). Again, the results from robust experiment and fallible observation are clearly non-compatible.

This litany of failure has attracted considerable popular comment. Medical journalist James Le Fanu has proposed an extreme solution to this problem: "The simple expedient of closing down most university departments of epidemiology could both extinguish this endlessly fertile source of anxiety mongering while simultaneously releasing funds for serious research."¹¹

Data dredging, biases, and confounding

It would seem wiser to attempt a better diagnosis of the problem before prescribing Le Fanu's solution. Data dredging is thought by some to be the major problem: epidemiologists have studies with a huge number of variables and can relate them to a large number of outcomes, with one in 20 of the associations examined being "statistically significant" and thus acceptable for publication in medical journals.¹² The misinterpretation of a $P < 0.05$ significance test as meaning that such findings will be spurious on only 1 in 20 occasions unfortunately continues. When a large number of associations can be looked at in a dataset where only a few real associations exist, a P value of 0.05 is compatible with the large majority of findings still being false positives.¹³ These false positive findings are the true products of data dredging, resulting from simply looking at too many possible associations. One solution here is to be much more stringent with "significance" levels, moving to $P < 0.001$ or beyond, rather than $P < 0.05$.¹⁴

Selection and information biases also need to be considered. Selection bias could produce a study database in which a given exposure is related to a variety of characteristics that increase (or decrease) risk of disease, where such associations are not apparent in the general population. Information biases also arise. For example, some people like to complain and will, if asked, complain both about life's experiences (such as stress) and also subjective health outcomes (such as having chest pain). An association between the two would lead to the inference that life stressors lead to angina, but in fact the two are simply related by a proclivity to complain, as evidenced by the finding that there is no association between reporting life stressors



Extra figures and references appear on bmj.com

BMJ 2002;325:1437-8

and objective, as opposed to subjective, indicators of coronary heart disease.¹⁰

By far the most likely cause of spurious association is confounding—where one factor that is not itself causally related to disease is associated with a range of other factors that do increase disease risk. Women who use hormone replacement therapy may be less likely to be smokers, more likely to exercise regularly, and less likely to be poor, all of which reduce the risk of coronary heart disease (see fig B on bmj.com). Associations reported in observational studies but not confirmed in randomised controlled trials tend to be of exposures that are related to many socioeconomic and behavioural measures that are in turn related to disease. As with bias, increasing the significance level provides no protection against being misled by confounded associations.

The inadequately recognised truth is that we live in an associational world—people who are disadvantaged in one regard tend to be disadvantaged in other regards, since the forces that structure life chances and experience tend to ensure that some folk get the worst of all things. We showed this by producing a pairwise correlation matrix of 133 physical examination and laboratory assay variables (8778 correlations) derived from a study of over 4000 older women.⁹⁸ This would be expected to yield 88 “significant” chance associations at the $P < 0.01$ level. In fact over 3000 such correlations were observed with a P value < 0.01 . In many ways it is more remarkable when things don’t “significantly” correlate with each other than when they do.

A standard argument is that hypotheses built on good scientific understanding of pathogenesis are unlikely to be spurious, but unfortunately it is generally easy to find a biologically plausible mechanism to “explain” each association.⁹⁹ Furthermore, it is seldom recognised how poorly the standard statistical techniques “control” for confounding, given the limited range of confounders measured in many studies and the inevitable substantial degree of measurement error in assessing the potential confounders.^{99 w10}

What can be done about confounding?

Where possible, associations should be replicated in databases in which the potential confounding structure differs from the initial study. In different countries exposures such as self reported stress, diet, or birth dimensions, for example, may be related in different ways to socioeconomic circumstances and socio-economically patterned causes of disease. Finding the same association within different populations gives some protection against being misled by confounding.

Specificity of associations between exposure and diseases is also helpful, as most diseases have only a finite number of causes. When exposures are related in a promiscuous way with a wide variety of outcomes, confounding by socially patterned behavioural and environmental factors is likely. Early on in the hormone replacement therapy debate, Diana Petitti pointed out that hormone replacement therapy apparently protected against accidental and violent deaths in observational studies as much as against coronary heart disease—and that given the lack of a plausible biological link between hormone replacement therapy and accidental or violent death, both associations may have been confounded.¹¹ This suggestion was later confirmed by the randomised controlled trials.⁴

Further measures include improving study design by measuring confounders better and thus allowing for a greater degree of statistical control. This may require carrying out more measurements on a smaller number of participants.^{w11} Sensitivity analyses should be carried out to model the degree to which measurement error in confounders could have left residual confounding^{w12 w13} and should be a necessary part of the statistical reporting of study results. A gift to epidemiology from modern genomics is the potential for using functional genetic polymorphisms that mimic the effects of environmental exposures to test exposure-disease relationships. There is very little opportunity when alleles segregate—effectively a random process—for social and behavioural factors to confound the resulting polymorphism-disease associations.^{12 w14}

Finally, the findings in observational studies of individuals should be related to the differences in risk of disease observed between populations, and within populations over time, as only those exposures which fit coherently into this scheme are likely to be important causes of disease.

Of course all our recommendations should be suspended once a year, to allow the Christmas issue of the *BMJ* to continue with its tradition of making the festive time a merrily data dredged, biased, and confounded one. Also remember that dredging, now disparaged, was the technique by which pearls were harvested from oysters. Among data dredged observations will reside new and precious associations: the only problem is deciding which ones should be gathered and used.

George Davey Smith *professor of clinical epidemiology*
(zetkin@bristol.ac.uk)

Shah Ebrahim *professor in epidemiology of ageing*

Department of Social Medicine, University of Bristol, Bristol BS8 2PR

Competing interests: GDS and SE are the co-editors of the *International Journal of Epidemiology*. Because the *BMJ* and other major weekly medical journals have cornered the market in splashing data dredged, biased, and confounded associations across the media through their press releases, the profile of quality journals is reduced, much to the chagrin of their editors.

- 1 Band PR, Le ND, Fang R, Deschamps M. Carcinogenic and endocrine disrupting effects of cigarette smoke and risk of breast cancer. *Lancet* 2002;360:1044-9.
- 2 Collaborative Group on Hormonal Factors in Breast Cancer. Alcohol, tobacco and breast cancer—collaborative reanalysis of individual data from 53 epidemiological studies, including 58 515 women with breast cancer and 95 067 women without the disease. *Br J Cancer* 2002;87:1234-45.
- 3 Stampfer MJ, Colditz GA. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med* 1991;20:47-63.
- 4 Beral V, Banks E, Reeves G. Evidence from randomised trials on the long-term effects of hormone replacement therapy. *Lancet* 2002;360:942-4.
- 5 Peto R, Doll R, Buckley JD, Sporn MB. Can dietary beta-carotene materially reduce human cancer rates? *Nature* 1981;290:201-8.
- 6 Willett WC. Vitamin A and lung cancer. *Nutrition Rev* 1990;48:201-11.
- 7 Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med* 1994;330:1029-35.
- 8 Khaw K-T, Bingham S, Welch A, Luben R, Wareham N, Oakes S, et al. Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study. *Lancet* 2001;357:657-63.
- 9 Heart Protection Study Collaborative Group. MRC/BHF heart protection study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002;360:23-33.
- 10 Macleod J, Davey Smith G, Heslop P, Metcalfe C, Carroll D, Hart C, et al. Psychological stress and cardiovascular disease: empirical demonstration of bias in a prospective observational study of Scottish men. *BMJ* 2002;324:1247-51.
- 11 Petitti DB, Perlman JA, Sidney S. Postmenopausal estrogen use and heart disease. *N Engl J Med* 1986;315:131-2.
- 12 Davey Smith G, Ebrahim S. Mendelian randomisation: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* (in press).