passed a child health insurance bill, and we had to stop enrolling children because we ran out of money. It was only at the very last moment, at this last special legislative session, through extraordinary efforts by some really committed folks in the governor's office, that a deal was struck to expand the programme so we could reopen enrolment. The programme provides subsidised health insurance to children who do not qualify for Medicaid for primary ambulatory care—not hospital care, no medications, no high-tech procedures, just basic routine care, getting the infection treated before it turns into a major problem. It's a very narrow, circumscribed programme. But we need to expand it to all eligible children. The legislators chose children because frankly, from a political standpoint, children are easiest to sell. But as we discussed, there's so much need for expanded access to programmes like this, even before we get to comprehensive reform. It can be effective, I think, in targeting the most needy populations and solving some of the biggest inequities in the system. So I think we can push ahead with things like insurance reform as we did in this legislative session—community rating, eliminating pre-existing conditions. We can push ahead with patchwork programmes like child health. Any amount of resources we can commit to that while we're trying to deal with the larger issues—how we're going to control costs in the major programmes and expand access and make sure that quality doesn't suffer at the same time. We have to do both.

So I'd look forward to a lot of beneficial changes. I'd look forward to a more enlightened AIDS policy if we get a change of administration in Washington. I would hope even if we don't get a change of administration we would get a more enlightened AIDS policy, but I don't really see much prospect of that.

1 Chassin MR. Quality of care: time to act. *JAMA* 1991;266:472-3.

Confounding in epidemiological studies: why "independent" effects may not be all they seem

George Davey Smith, Andrew N Phillips

> Nobody mentions that the children of parents conscientious and careful enough to have their children "immunized" will come out of any statistical test better than the children of the comparatively careless. Poverty, too, produces startling vital statistics which can be turned to account by the exploiters of any nostrum. If the jewellers had thought of claiming that the possession of a gold watch and chain is an infallible prophylactic against smallpox, their statistics would have been quite as convincing as those of the vaccinists.[1]

In this quote George Bernard Shaw deals with an issue central to the interpretation of epidemiological studies: how is it possible to decide whether a factor is causally related to a health outcome, rather than simply being associated with factors which are truly causal? In epidemiological parlance, the issue at stake is that of confounding. This is illustrated in figure 1. Imagine that exposure A refers to smoking and exposure B to yellow fingers and the outcome is lung cancer. Through being associated with smoking, yellow fingers will be related to lung cancer risk. In this case interpretation is easy, yet how in general do we separate such spurious associations from ones that might be causal?

The issue is of more than parochial interest, since general medical journals devote much of their space to publishing the results of observational epidemiological studies that examine whether there are health risks associated with a particular exposure. It is not surprising, perhaps, that these reports attract considerable media attention—especially when they show apparent hazards consequent on aspects of daily life.[2] Lately, for example, we have read that oral contraceptives facilitate the acquisition of HIV infection[3]; that coronary heart disease risk seems to be increased by drinking coffee,[4] not drinking alcohol,[5] allowing your teeth to rot[6] or having had a low birth weight[7]; that sloth predisposes to diabetes,[8] that not having received breast milk results in low intelligence,[9] and that smoking is a cause of cervical cancer.[10] Regarding the first of these apparently hazardous activities, however, a different group of investigators have more recently informed us that use of oral contraceptives, far from facilitating transmission of HIV, actually protects against the virus.[11] In this article we suggest that the phenomena bringing about these contradictory findings may distort many of the epidemiological associations that have created excitement regarding the possible identification of factors involved in disease aetiology.

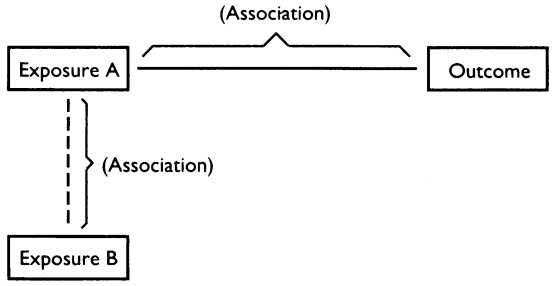

FIG 1—*Exposure A (for example, smoking) is associated with the outcome (lung cancer risk); exposure B (yellow fingers) is associated with exposure A—and hence to the outcome*

Confounding in practice

Epidemiological studies of cervical cancer have, over the years, identified a myriad of risk factors for the disease; these have included not eating carrots, a history of induced abortion, drinking alcohol, practising masturbation at an early age, low dietary intake of folate, use of oral contraceptives, and high parity. That a sexually transmissible agent is involved is now generally accepted, but much attention has also been given to cigarette smoking, with a large series of studies having reported that smoking is associated with the risk of cervical cancer.[12]

If smoking is related to the risk of exposure to the sexually transmissible agent, then the association between smoking and cervical cancer could be due to confounding by this agent: following the scheme of the figure, exposure A would be the sexually transmitted agent and exposure B would be cigarette smoking. Much data indicate that this possibility should be taken seriously. Cigarette smoking is, for example, strongly associated (odds ratio of 7·2) with early loss of

Department of Public Health, University of Glasgow, Glasgow G12 8RZ
George Davey Smith, *senior lecturer in public health*

Academic Department of Genitourinary Medicine, University College and Middlesex School of Medicine, London W1N 8AA
Andrew N Phillips, *lecturer in epidemiology*

Correspondence to:
Dr Davey Smith.

BMJ 1992;305:757-9

HULTON DEUTSCH

FIG 2—*Epidemiological studies have identified smoking—including passive smoking—as a risk factor for cervical cancer*

virginity[13] and with higher numbers of sexual partners during early adulthood,[14] both of which would be related to the risk of acquiring a sexually transmitted infectious agent.

Confounding similar to that suggested between smoking and exposure to a sexually transmitted agent as putative risk factors for cervical cancer could occur for the other associations mentioned above. Thus, lack of physical activity is related to obesity; as obesity is a powerful risk factor for diabetes in its own right the relation between lack of physical activity and diabetes could be due to confounding by obesity. Similarly, the practice of providing breast milk is associated with a host of social variables that could determine a child's ability to score highly on an IQ test.

Dealing with confounding

The possibility that confounding could account for observed associations in epidemiological studies is generally dealt with through the statistical demonstration that the associations are "independent" of the confounding factors.

The usual approach when seeking to establish the "independence" of associations is to fit a multiple regression model. When the occurrence of disease or death is the outcome of interest, the logistic model or the Cox proportional hazards model is generally used. An exposure is said to be independently associated with the outcome if the association remains after the levels of other exposures are held fixed—commonly termed "controlling" or "adjusting" for other exposures. The statement that there is an association between physical activity and risk of diabetes which is independent of obesity means that among subjects with an equal degree of obesity those who are less physically active are more likely to develop diabetes than those who are more active. As a further example, consider the first situation discussed in relation to figure 1. Smoking is associated with lung cancer risk. Yellow fingers are also associated with lung cancer risk, but only because they are associated with smoking. Because the association between yellow fingers and lung cancer risk acts through smoking, the association between yellow fingers and lung cancer risk would disappear after the effect of smoking has been controlled for. In contrast, smoking would remain associated with lung cancer risk after the presence of yellow fingers has been controlled for. Thus, in theory, these methods can be used to identify exposures that are independently related to risk of the outcome and therefore possibly causal.

Limitations of this approach

The use of the approach outlined can, however, produce misleading results. In particular, a situation could arise in which an exposure seems to have an association with risk of disease that is "independent" of another exposure when this is not in fact the case. This arises owing to the fact that when measuring an exposure we do not always characterise it accurately. Take, for example, systolic blood pressure. We might measure this at a baseline screen in a group of subjects who we plan to follow up to assess the incidence of stroke. Systolic blood pressure is highly variable: when it is measured on different occasions, different levels are obtained. In epidemiological studies, however, single readings of blood pressure are often used. Many individuals will be characterised incorrectly in this situation. If we assume that the "usual" level of systolic blood pressure, experienced over a long period of time, is the most important aspect of blood pressure in determining risk of stroke, then this is what should be measured. This, however, is difficult to do in large numbers of subjects. If only one measure in each subject is taken then the exposure we have measured and called "systolic blood pressure" is an imprecise measure of what we really want to know.

Now imagine that in the same baseline screen in which blood pressure was measured, study subjects were given an electrocardiograph and they were assessed for left ventricular hypertrophy, a condition that frequently develops in people who suffer from hypertension. If we find when analysing the results of the study that left ventricular hypertrophy as well as systolic blood pressure is apparently independently associated with risk of stroke, as some studies have suggested,[15,16] we may be being misled. This is because left ventricular hypertrophy serves as a marker of usual systolic blood pressure, albeit an even cruder one than a one off, direct measurement of blood pressure. Each measure is therefore able to provide some information about subjects' usual systolic blood pressure (and hence their risk of stroke) over and above that provided by the other. Thus the association between left ventricular hypertrophy and risk of stroke may remain after adjustment for systolic blood pressure only because of the inadequacy of a single measure of systolic blood pressure as a measure of usual systolic blood pressure, and nothing more. In this situation, the association between left ventricular hypertrophy and risk of stroke after adjustment for systolic blood pressure would be said to be due to residual confounding, a phenomenon the importance of which is often underappreciated.[17,18]

In the case of smoking and cervical cancer risk, responses to questions regarding sexual behaviour are used as proxy measures of risk of exposure to a sexually transmitted agent which may be involved in the aetiology of the disease. The association between cervical cancer and smoking is examined after adjusting for reported sexual behaviour. The independent association that remains between cigarette smoking and cervical cancer after such adjustment may be due to the inadequacy of questions concerning sexual activity as a measure of exposure to the sexually transmitted agent. Similarly, the independent association between provision of breast milk and child's IQ[9] may arise because of the inadequacy of father's job and mother's education as measures of the social

factors that might determine a child's ability to score highly on intelligence tests.[19]

Finally, in the study that reported an independent association between physical activity and risk of diabetes after controlling for body mass index,[8] body mass index was based on self reported height and weight. Given the strong associations between body mass index and both physical activity and diabetes, even the small amount of imprecision in assessment of body mass index could be enough to produce a spurious independent effect of physical activity. This possibility is supported by the finding that the increased risk of diabetes associated with low physical activity is noticeably attenuated by adjustment for body mass index, even when this is poorly characterised by the self reports of height and weight.

Biological plausibility

Biological plausibility is often appealed to in the interpretation of epidemiological associations.[20] For example, when discussing early masturbation as a risk factor for cervical cancer Rotkin recognised that "a reasonable physiological rationale" was required in order to "warrant expansion of research" in the directions his study indicated.[21] This rationale existed in the pathways from the nervous system to cervical tissue, stimulation of which could promote cancer. It is, perhaps, too easy to generate arguments regarding the biological plausibility of epidemiological findings. When the diametrically opposed associations between oral contraceptive use and the risk of acquiring HIV were identified, both research teams produced highly plausible explanations for their findings.[3 11] It is the exceptional epidemiological association for which no mechanism can be proposed, with many associations being safeguarded by not being obviously implausible.

Conclusions

It is likely that many of the associations identified in epidemiological studies are due to confounding, often by factors which are difficult to measure. Unfortunately, investigators often have a relatively cavalier approach to this possibility. The result is the generation of a multitude of risk factors for ill health,[22] which are often greeted with substantial publicity, followed by understandable public scepticism.[23] The first responsibility of investigators should be to retain the proper degree of caution when interpreting and discussing their findings.

A second response relates to the design of studies. Many epidemiological associations are detected through the analysis of data sets that were not established to investigate the issue at hand. In such situations it is likely that data will not have been gathered on all, if any, relevant confounding factors. This is one reason why what can be termed "unplanned observations" are highly susceptible to the production of spurious associations.[24] Even when studies are planned to investigate a particular issue, the problems of confounding and, especially, residual confounding are underappreciated. The optimal strategy is to perform studies in populations in which the exposure of interest and potential confounding factors are not associated with each other.

This is often not possible, however, and as discussed above, poor measurement of potential confounding factors can lead to there being little possibility of properly accounting for these factors during the analysis stage.[25-27] There is, however, a general tradition of designing studies that maximise the number of subjects, who are examined only once. This strategy may be misguided: the identification of epidemiological associations could often be better served by studies that measure fewer subjects more than once and classify both the exposure of interest and potential confounders more accurately.

1 Shaw GB. *Everybody's political what's what?* London: Constable, 1944.
2 Feinstein AR. Scientific standards in epidemiological studies of the menace of daily life. *Science* 1988;**242**:1257-63.
3 Plummer FA, Simonsen JN, Cameron DW, Ndinya-Achola JO, Kreiss JK, Gakinya MN, *et al.* Cofactors in male-female sexual transmission of human immunodeficiency virus type 1. *J Infect Dis* 1991;**163**:233-9.
4 Rosenberg L, Palmer JR, Kelly JP, Kaufman DW, Shapiro S. Coffee drinking and nonfatal myocardial infarction in men under 55 years of age. *Am J Epidemiol* 1988;**128**:570-8.
5 Rimm EB, Giovannucci EL, Willett WC, Colditz GA, Ascherio A, Rosner B, *et al.* Prospective study of alcohol consumption and risk of coronary disease in man. *Lancet* 1991;**338**:464-8.
6 Mattila KJ, Nieminen MS, Valtonen VV, Rasi VP, Kesaniemi YA, Syrjala SL, *et al.* Association between dental health and myocardial infarction. *BMJ* 1989;**298**:779-82.
7 Barker DJP, Winter PD, Osmond C, Margetts B, Simmonds SJ. Weight in infancy and death from ischaemic heart disease. *Lancet* 1989;ii:577-80.
8 Manson JE, Rimm EB, Stampfer MJ, Colditz GA, Willett WC, Krolewski AS, *et al.* Physical activity and incidence of non-insulin dependent diabetes mellitus in women. *Lancet* 1991;**338**:774-8.
9 Lucas A, Morley R, Cole TJ, Lister G, Leeson-Payne C. Breast milk and subsequent intelligence quotient in children born preterm. *Lancet* 1992;**339**: 261-4.
10 Slattery ML, Robison LM, Schuman KL, French TK, Abbott TM, Overall JC, *et al.* Cigarette smoking and exposure to passive smoke are risk factors for cervical cancer. *JAMA* 1989;**261**:1593-8.
11 Lazzarin A, Saracco A, Musicco M, Nicolosi A. Man-to-woman sexual transmission of the human immunodeficiency virus. *Arch Intern Med* 1991; **151**:2411-6.
12 Winkelstein W. Smoking and cervical cancer—current status: a review. *Am J Epidemiol* 1990;**131**:945-57.
13 Orr DP, Beiter M, Ingersoll G. Premature sexual activity as an indicator of psychosocial risk. *Pediatrics* 1991;**87**:141-7.
14 MacDonald NE, Wells GA, Fisher WA, Warren WK, King MA, Doherty JA, *et al.* High-risk STD/HIV behavior among college students. *JAMA* 1990;**263**:3155-9.
15 Kannel WB. Prevalence and natural history of electrocardiographic left ventricular hypertrophy. *Am J Med* 1983;**75**:4-11.
16 Shaper AG, Phillips AN, Pocock SJ, Walker M, Macfarlane PW. Risk factors for stroke in middle aged British men. *BMJ* 1991;**302**:1111-5.
17 Davey Smith G, Phillips AN. Declaring independence: why we should be cautious. *J Epidemiol Community Health* 1990;**44**:257-8.
18 Savitz DA, Baron AE. Estimating and correcting for confounder misclassification. *Am J Epidemiol* 1989;**129**:1062-71.
19 Doyle LW, Rickards AL, Kelly EA, Ford GW, Callanan C. Breastfeeding and intelligence. *Lancet* 1992;**339**:744-5.
20 Angell M. The interpretation of epidemiological studies. *N Engl J Med* 1990;**323**:823-5.
21 Rotkin ID. Sexual characteristics of a cervical cancer population. *Am J Public Health* 1967;**57**:815-29.
22 Hopkins PN, Williams RR. A survey of 246 suggested coronary risk factors. *Atherosclerosis* 1981;**40**:1-52.
23 Frankel S, Davison C, Davey Smith G. Lay epidemiology and the rationality of responses to health education. *Br J Gen Pract* 1991;**41**:428-30.
24 Bradford Hill A. Observation and experiment. *N Engl J Med* 1953;**248**: 995-1001.
25 Phillips AN, Davey Smith G. How independent are "independent" effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol* 1991;**44**:1223-31.
26 Phillips AN, Davey Smith G. Bias in relative odds estimation due to imprecise measurement of correlated exposures. *Statistics in Medicine* 1992;**11**:953-61.
27 Measurement imprecision: Ignore or investigate? *Lancet* 1992;**339**:587-8.

(Accepted 10 July 1992)