Review

Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies

Stephen MacMahon, Rory Collins

Observational studies and randomised trials can contribute complementary evidence about the effects of treatment on mortality and on major non-fatal outcomes. In particular, observational studies have an important role in the identification of large adverse effects of treatment on infrequent outcomes (ie, rare, but serious, side-effects) that are not likely to be related to the indications for (or contraindications to) the treatment of interest. Such studies can also provide useful information about the risks of death and disability in particular circumstances that can help to generalise from clinical trials to clinical practice. But, due to their inherent potential for moderate or large biases, observational studies have little role in the direct assessment of any moderate effects of treatment on major outcomes that might exist. Instead, sufficiently large-scale evidence from randomised trials is needed to assess such treatment effects appropriately reliably. Wider appreciation of the different strengths and weaknesses of these two types of epidemiological study should increase the likelihood that the most reliable evidence available informs decisions about the treatments doctors use—and patients receive—for the management of a wide range of life-threatening conditions.

Epidemiological studies of the effects of treatments on mortality and major non-fatal outcomes can take the form of either clinical trials or observational studies. The first part of this review[1] dealt with clinical trials—in particular, those in which the treatment is assigned to patients at random. As discussed, randomisation minimises systematic errors (ie, biases) in the estimates of treatment effects, allowing any moderate effects that exist to be detected unbiasedly in studies of appropriately large size.[1] By contrast, observational studies—such as cohort studies and case-control studies—involve comparisons of outcome among patients who have been exposed to the treatment of interest, typically as part of their medical care, with outcome among others who were not exposed (or comparisons between those with different amounts of exposure). The reasons why certain patients received a particular treatment while others did not are often difficult to account for fully, and, largely as a consequence, observational studies are more prone to bias than are randomised trials. The primary objective of the second part of this review is to distinguish between situations in which biases in observational studies could lead to misleading conclusions and those in which such studies could provide useful evidence about the effects of treatment.

OBSERVATIONAL STUDIES: Non-randomised assessment of treatment

Assessment of adverse effects of treatment

Observational studies can have an important role in the identification of large adverse effects of treatments, particularly on infrequent outcomes that are not likely to be related to the indications for, or contraindications to, the treatment of interest (panel 1). Perhaps one of the best illustrations of this is the detection of increased risks of abnormal fetal limb development after maternal use of thalidomide.[2] A decade later, observational studies also detected the many-fold increased risk of vaginal clear-cell adenocarcinoma among the daughters of women who used diethylstilboestrol.[3] Other more recent examples include the demonstration of a 20-fold increased risk of cardiac-valve abnormalities among patients taking the appetite-suppressant drugs fenfluramine, dexfenfluramine, and phentermine[4] (table 1), and even larger increases in the risk of Stevens-Johnson syndrome and toxic epidermal necrolysis with antiepileptic therapy.[5] In each of these examples, the outcome was rare among unexposed individuals and the excess risk was large among exposed individuals, making it unlikely that systematic errors could reasonably account for the entire association.

On the other hand, since the disease of interest is rare in such circumstances, individual studies may well involve too few cases to detect, or quantify reliably, even large increases in risk. Hence, to minimise random error, combined analyses of the aggregated results (ie, meta-analyses) of all relevant observational studies are being done with increasing frequency. For example, a meta-analysis found that more than 10 years of oestrogen replacement therapy unopposed by progestagen was associated with almost a ten-fold increase in the risk of endometrial cancer among postmenopausal women.[6] Such large effects are unlikely to be entirely the consequence of bias, but it is not so easy to exclude the possibility that biases might largely or wholly explain more modest increases in risk: for

Panel 1: **Situations in which an observational study is more likely to provide reliable evidence about adverse effects of treatment**

- The outcome of interest is rare among individuals not exposed to the treatment
- The excess risk among individuals exposed to the treatment is large (eg, a several-fold increase in risk)
- There are no obvious sources of bias likely to account for most, or all, of the observed association

Lancet 2001; **357:** 455–62

Institute for International Health, University of Sydney, PO Box 576, Newtown, Sydney, New South Wales 2042, Australia (Prof S MacMahon PhD)**; and Clinical Trial Service Unit & Epidemiological Studies Unit, University of Oxford, Oxford, UK** (Prof R Collins MBBS)

Correspondence to: Prof Stephen MacMahon (e-mail: s.macmahon@iih.usyd.edu.au)

	Any appetite suppressant (n=233)*	Unexposed controls (n=233)†	Odds ratio (95% CI)
Valve abnormalities	53 (23%)	3 (1%)	22·6 (7·1–114·2)

*163 on fenfluramine and phentermine, 31 on dexfenfluramine and phentermine, and 39 on dexfenfluramine alone. †Matched for sex, age, height, and body-mass index.

Table 1: **Detection of large adverse effects of treament in an observational study: cardiac-valve regurgitation with appetite-suppressant drugs[4]**

example, the 40% increased incidence of malignant melanoma seen among users of hormone replacement therapy in another meta-analysis of observational studies.[7] For, although these meta-analyses may help avoid the biases produced by unduly selective emphasis on particular parts of the available evidence (as with meta-analyses of randomised trials[1]), the combination of observational evidence that is subject to other systematic errors might merely compound those biases—that is, produce more precise, but still biased, estimates of the effects of treatment.

Assessment of beneficial effects of treatment

Reliable evidence about the effects of treatment on mortality and major morbidity can also emerge from observational studies when outcome among untreated patients is typically poor and a large proportion of patients derive benefit from the treatment. For example, the beneficial effects of penicillin on survival in patients with sepsis,[8] and of antihypertensive treatment on death and stroke in patients with malignant hypertension,[9] were demonstrated in simple case series. Another example is provided by oral rehydration therapy, which seemed to reduce mortality from about 30% to less than 5% when introduced during a cholera epidemic among Bangladeshi refugees.[10] But, as discussed in the first part of this review,[1] outcome for many other common serious conditions is less predictable, and the most plausible expectation of benefit is that a treatment produces only moderate (although still potentially worthwhile) effects on serious outcomes. For the reliable assessment of such effects of treatment, observational studies have a much more limited part to play,[11] since the potential biases could obscure, inflate, or even seem to reverse the real effects of treatment—and these biases cannot be quantified reliably. For example, whereas a case-control study indicated that individuals treated with angiotensin-converting-enzyme inhibitors had a 30% lower risk of cancer,[12] such effects have not been confirmed by large randomised trials of these agents.[13–15] Hence, although observational studies have found hormone replacement therapy to be associated with a third less coronary heart disease[16] and a fifth less colorectal cancer[7] among postmenopausal women, the extent to which such differences in risk reflect treatment effects rather than biases (due, perhaps, to the preferential prescription of hormone replacement therapy to lower-risk women) may only be known when results emerge from large randomised trials.[17] Similarly, the observation in a case-control study of a 70% lower risk of dementia among individuals treated with statins cannot be accepted as good evidence of benefit from the treatment.[18]

Major sources of bias in observational studies

Confounding by factors associated with both treatment and outcome

Perhaps the most important potential source of bias in observational studies is confounding, whereby some factor is associated with the exposure of interest—but is not a direct consequence of it—and, independently, influences the risk of the outcome of interest (panel 2). Observational studies of the effects of exposure to treatment are particularly prone to confounding by indication (or by contraindication), with the development of a medical condition leading both to the use of the treatment (or its avoidance) and to the outcome of interest. This type of bias can produce misleading estimates not just of the size but also of the direction of treatment effects, depending on the nature of the associations between the confounding factors and the outcome.

Panel 2: **Major sources of bias in observational studies of treatment**

- ***Confounding:*** A factor (such as pre-existing disease severity) is associated with the use (or avoidance) of the treatment and, independently, influences the risk of the outcome of interest.

  For example, "confounding by indication (or contraindication)" may occur when the treatment tends to be provided more (or less) frequently to individuals with medical conditions associated with increased or decreased risks of the outcome of interest
- ***Recall bias:*** The reliability of recall of treatment exposure differs between those who develop an adverse outcome and those who do not
- ***Detection bias:*** The reliability of detection of adverse outcomes differs between those exposed to the treatment of interest and those not exposed

A recent example of misleading evidence about the size of a treatment effect is provided by a large observational study in which patients who received β-blockers after myocardial infarction were about half as likely to die as those who did not receive such treatment (table 2).[19] By contrast, large-scale evidence from randomised trials has clearly shown that long-term β-blocker use in patients with a history of myocardial infarction reduces the risk of death by only about a quarter[20] (as have trials in higher-risk patients with congestive heart failure[21–23]). The patients who received β-blockers in this observational study were significantly younger, and had a lower-risk medical history, than those who did not. Statistical adjustments were made for these, and other, potential confounding factors that had been recorded, but such adjustments may well be incomplete due both to insufficient correction for factors that were recorded (because of random errors in their measurement[24]) and to lack of correction for other relevant factors. Hence, it seems likely that the overestimation in this observational study of the survival advantage produced by β-blocker therapy reflects some residual bias (due, perhaps, to a selective tendency for these drugs to be used less frequently in higher-risk patients).

	Deaths/patients		Risk ratio† (95% CI)
	β-blocker*	No β-blocker*	
Observational study‡	~123/785 (16%)	~886/2952 (30%)	0·57 (0·47–0·69)
Randomised trials	827/10 452 (8%)	986/9860 (10%)	0·77 (0·70–0·85)

*Treatment recorded at baseline in the observational study and assigned at random in the trials. †Multivariate adjusted relative risk in the observational study, and stratified odds ratio in the meta-analysis of randomised trials. ‡Exact numbers in each treatment group of the observational study were not reported.

Table 2: **Different sizes of apparent effect in an observational study[19] and in randomised trials:[20] β-blocker use and death after myocardial infarction**

	CHD events/patients		Risk ratio† (95% CI)
	Antihypertensive therapy*	No antihypertensive therapy*	
Observational study	50/839 (6%)	420/20 475 (2%)	1·8 (1·3–2·6)
Randomised trials	934/23 847 (4%)	1104/23 806 (5%)	0·84 (0·77–0·92)

CHD=coronary heart disease. *Treatment recorded at baseline in the observational study and assigned at random in the trials. †Multivariate adjusted relative risk in the observational study and stratified odds ratio in the meta-analysis of randomised trials.

Table 3: **Different directions of apparent effect in an observational study[25] and in randomised trials:[26] antihypertensive therapy and coronary heart disease**

	Stroke/patients		Risk ratio† (95%CI)
	Lower-dose aspirin (<650 mg daily)*	Higher-dose aspirin (650–1300 mg daily)*	
Observational study	96/1391 (7%)	15/835 (2%)	2·3 (1·3–3·9)
Randomised trial	64/1417 (5%)	86/1432 (6%)	0·74 (0·53–1·03)

*Treatment recorded at baseline in the observational study and assigned at random in the randomised trial. †Univariate relative risk in the observational study and odds ratio in the randomised trial.

Table 4: **Discordance between apparent effects of different drug doses in an observational study[29] and a randomised trial:[30] higher-dose versus lower-dose aspirin and stroke after carotid endarterectomy**

An example of misleading evidence about the direction of a treatment effect is provided by an observational study in which there was almost a two-fold greater risk of coronary events among patients receiving antihypertensive therapy than among those not receiving such treatment (table 3).[25] By contrast, randomised controlled trials have clearly demonstrated that antihypertensive treatment reduces the risks of coronary heart disease (as well as those of stroke).[26] Similarly, whereas a large observational study found nearly a doubling in the risk of major coronary events among those regularly taking aspirin,[27] randomised controlled trials have shown unequivocally that antiplatelet therapy reduces the risks of heart attacks by about a quarter.[28] These misleading findings from observational studies persisted after statistical adjustment for a variety of confounding factors and after restriction of analyses to individuals without a recorded history of cardiovascular disease. Once again it seems that uncontrolled residual bias remains the most likely explanation (probably, in these examples, due to a tendency for the treatments to be used more frequently in higher-risk patients). Fortunately, for both antihypertensive and antiplatelet therapy, the evidence from the randomised trials has chiefly influenced practice patterns, resulting in the appropriately widespread use of these treatments and the consequent prevention of many hundreds of thousands of premature deaths each year. By contrast, reliance on the evidence from the observational studies might have led to the inappropriate abandonment of these treatments (or, at the very least, to restriction of their use) and to much unnecessary suffering.

Observational studies can also provide misleading evidence about the effects of different drug doses. For example, retrospective observational analyses of outcome among participants in the North American Symptomatic Carotid Artery Endarterectomy Trial (NASCET) indicated that the risk of perioperative stroke among patients who had been taking 650–1300 mg aspirin daily was less than half that among patients who had taken lower doses (table 4).[29] Subsequently, however, a randomised trial designed to test this hypothesis in patients undergoing carotid endarterectomy found a non-significantly higher stroke incidence with 650–1300 mg/day aspirin than with lower doses (as well as a marginally significant higher risk of the composite of stroke, myocardial infarction or death).[30] In this instance, reliance on the evidence from the observational study alone could have led to the inappropriate abandonment of lower-dose regimens which cause fewer side-effects and are better tolerated long-term.

Bias due to differential recall of treatment exposure

Recall bias can be a problem in observational studies when there is a difference in the reliability of the data collected on treatment exposure between cases that have the disease of interest and controls that do not.[31] Although it is unlikely that recall bias could account for the many-fold increases in risk seen, for example, with limb abnormalities and thalidomide use,[2] it might well be responsible for more moderate differences in apparent risk. For example, an early case-control study of childhood cancer obtained data on maternal X-ray exposure through interviews with mothers, and observed that the risk of death from malignancy among the children of women who reported being exposed to abdominal X-rays was almost twice as great as that among the children of women who reported no such exposure.[32] To determine whether this association might, at least in part, reflect more complete recall of exposure by the mothers of affected children, a second study was done in which exposure was determined from prenatal medical records.[33] That study also found an increased risk of cancer among offspring of exposed women, but the relative risk was only half as large as in the first study. It has been suggested that such bias might be kept to a minimum by making comparisons between exposures reported by mothers of children with some particular birth defect and those reported by mothers of children with other anomalies.[34] That strategy would not, however, exclude entirely the possibility of differential recall between the mothers of children with different types of birth defect. Moreover, it might obscure a real effect of the treatment if exposure was associated with more than one type of congenital anomaly.

Bias due to differential detection of outcomes

Individuals receiving any treatment will tend to be seen by doctors or other health professionals more frequently than will others, and this may result in the earlier detection of a variety of outcomes. For example, although a highly significant increase of a quarter in the risk of breast cancer was seen among women taking hormonal contraceptives,[35] this finding could largely reflect the earlier detection of less advanced breast cancer among such women. For, much of the observed excess risk was due to an excess of localised tumours, without any clear increase in the risk of tumours that had spread beyond the breast. Another possible example of such detection bias is provided by studies of first-trimester exposure to the antifungal drug itraconazole. Congenital malformations were seen in 13% of children of exposed women in a retrospective study compared with only 3% in a prospective study,[36] perhaps reflecting the greater likelihood of including women who have affected babies in a retrospective study.

Efforts to control biases in observational studies

The effects of biases in observational studies are frequently underestimated in the interpretation of

associations found between treatment and outcome. Even when statistical adjustment for measured confounding factors fails to reduce the size of such associations materially, this provides little reassurance that residual bias is not still a major cause of any observed associations. These difficulties are illustrated by an observational study of antihypertensive treatment in which a 60% higher risk of heart attacks was seen among patients receiving a calcium antagonist compared with those receiving other agents.[37] In that study, calcium antagonists seem to have been preferentially prescribed to higher-risk patients (such as those with pre-existing coronary heart disease or other risk factors for cardiovascular disease), but the association between use of calcium antagonists and subsequent myocardial infarction remained conventionally significant after adjustment for measured confounders and after excluding those with a history of cardiovascular disease. Residual bias remains a plausible explanation for at least part of the observed excess risk, however, since the data collected on prognostic factors are unlikely to describe all of the factors that contributed to the tendency to prescribe calcium antagonists to higher-risk patients. This could explain why the large excess risk reported in that observational study is not consistent with the much smaller difference in heart attack incidence (relative risk 1·12 [95% CI 1·00–1·26]) between calcium antagonists and diuretics or β-blockers in a prospectively-planned meta-analysis of all relevant randomised trials.[38]

Various statistical methods have been proposed to deal with the problem of residual biases in observational studies of treatment. For example, instrumental variable estimation involves grouping patients according to their likelihood of receiving the treatment of interest, by use of observable factors (ie, instrumental variables) that affect treatment use, but—it is hoped—do not directly affect patients' outcomes.[39] Although this method has been described as mimicking randomisation, it depends entirely on the untestable assumption that the observed instrumental variables are not correlated with unobserved factors that directly affect outcome. Moreover, since the range of variation between groups of patients in the likelihood of receiving some particular treatment might be narrow (eg, one such assessment of coronary-artery catheterisation was based on its use in 20% *vs* 26% of patients[39]), any difference in outcome due to this differential use of the treatment would probably be very small (and, hence, difficult to assess even in a properly randomised controlled trial).

Another method that has been proposed involves case-crossover (or case-series) analysis,[40,41] in which outcomes are compared between periods before and after treatment exposure within the same individuals. But, although this may avoid biases resulting from differences between exposed and non-exposed patients, variations in the underlying disease state within individuals could still determine both the necessity for treatment and the likelihood of the outcome of interest occurring. For example, a case-crossover study reported a 60% higher risk of road-traffic accidents during periods of exposure to benzodiazepines.[42] At least in part, this could have been due to exacerbation of certain conditions that led both to an increased use of benzodiazepines and, independently, to an increased risk of accidents. Hence, these and other non-randomised methods[43] do not provide assurance that all sources of known and unknown bias are adequately controlled, and so cannot exclude the possibility that moderate biases have obscured or inflated any moderate effects, or have falsely indicated a treatment effect when none existed.

Potential for small random errors in observational studies

One advantage of observational studies is that it is often easier to study much larger numbers of patients—and, consequently, much larger numbers of deaths and other relevant outcomes—than it is in randomised trials. Observational studies can, therefore, provide estimates of treatment effects that are subject to relatively small random errors, allowing the reliable detection of some extreme though rare adverse effects of treatments. But, as discussed earlier, small random errors in large observational studies can also lead to the detection of more moderate differences in risk that are merely the result of bias, rather than the effect of treatment (ie, more precise, but biased, estimates). For example, in meta-analyses of observational studies of hormone replacement therapy, women who had taken such treatment were seen to have significantly lower risks of coronary heart disease with oestrogen alone (relative risk 0·70 [95% CI 0·65–0·75]) or with oestrogen plus progestin (0·66 [0·53–0·84]),[16] lower risks of colorectal cancer (0·8 [0·7–0·9]),[7] higher risks of breast cancer (relative risk increasing by 2·3% [1·1–3·6] with each year of use),[44] and higher risks of malignant melanoma (relative risk 1·4 [1·2–1·7]).[7] But, there is evidence that women who take hormone replacement therapy may have better pretreatment coronary risk-factor profiles[45] and better access to preventive health care[46] than those who do not, and several of the risk factors for coronary heart disease that differ between users and non-users of hormone replacement therapy (such as physical inactivity and obesity) are also risk factors for colon cancer.[47] On the other hand, women who take hormone replacement therapy (like those who take oral contraceptives[35]) may be more likely to have breast cancer and melanoma diagnosed at an earlier stage because of greater contact with doctors. As a consequence, the balance of any true benefits and risks of hormone replacement therapy cannot be determined reliably from observational studies. In this regard, it is of interest that a relatively small randomised placebo-controlled trial of hormone replacement therapy for the secondary prevention of coronary heart disease[48] failed to confirm the one-third reduction in risk suggested by the observational studies.[16] Indeed, during about 4 years of follow-up in that study, there seemed to be an early excess of vascular events followed by a later shortfall among those assigned active treatment (and a similar early trend has recently been reported from the larger Women's Health Initiative randomised trial[49]). These findings do not, however, preclude the possible emergence of worthwhile effects with more prolonged use of hormone replacement therapy in the large, long-term trials that are currently in progress.[17]

Evidence from observational studies in the context of results from randomised trials

A more prominent role for observational studies in the assessment of treatment effects has been argued in two reviews[50,51] on the basis of examples in which there were considered to be no apparent differences between the results of observational studies and those of randomised trials. But, several of the examples included in those reviews involved estimates of treatment effects that were subject to large random errors. For example, separate meta-analyses of the observational studies and of the

randomised trials comparing laparoscopic and open appendectomy were interpreted as having shown similar reductions in infection rates with laparoscopic procedures,[50] even though the 95% CI for the risk reduction in each type of study ranged from about 10% to about 70%. It has also been noted[52] that other examples in those reviews did not even involve observational studies of treatment, but instead misleadingly compared the effects found in randomised trials of treatment that alter risk factors (such as lowering blood cholesterol or blood pressure) with estimates from observational studies of the associations between risk-factor levels and disease risk.[51] Most pertinently, any similarity of the treatment effects estimated from observational studies and from randomised trials in any one particular circumstance provides little reassurance that observational studies will provide unbiased estimates of the effects of treatment in some other circumstance.

Furthermore, it makes little sense to continue to base inference on observational studies when their results have been reliably refuted by large-scale randomised trials. For example, it had previously been suggested by observational analyses that digoxin might increase mortality substantially.[53] By contrast, the large randomised Digitalis Investigation Group (DIG) trial showed unequivocally that the addition of digoxin to current therapy reduces the risk of hospital admission for heart failure (relative risk 0·72 [95% CI 0·66–0·79]) by about as much as do angiotensin-converting-enzyme inhibitors, with no apparent adverse effect on total mortality (1181 deaths in patients allocated digoxin and 1194 in patients allocated placebo; 0·99 [0·91–1·07]).[54] In the light of this evidence, the prominent reporting of new claims from observational analyses that digoxin doubles the risk of death in just the sort of patients studied in the DIG trial is not appropriate.[55,56]

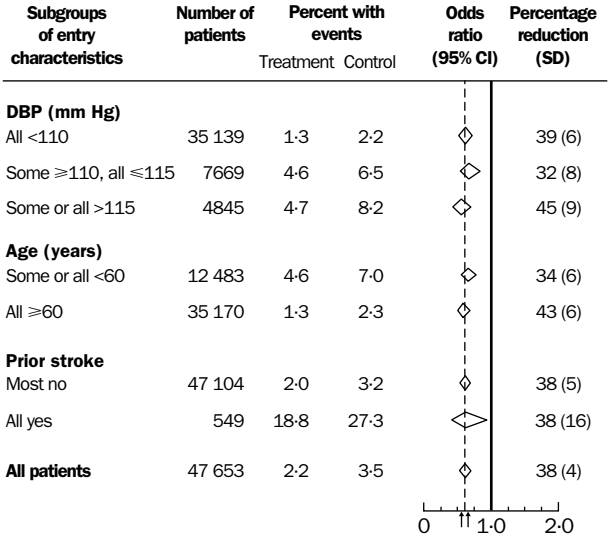
Without clear confirmatory evidence from large-scale randomised trials or their meta-analyses, reports of moderate treatment effects from observational studies should not be interpreted as providing good evidence of either adverse or protective effects of these agents (and, contrary to other suggestions,[57,58] the absence of evidence from randomised trials does not in itself provide sufficient justification for relying on observational data). In this regard, it is salutary to note the example provided by early reports from observational studies of moderately increased risks of breast cancer among hypertensive patients treated with reserpine.[59,60] Those reports led to avoidance of one of the few effective antihypertensive agents available at that time, and only much later was this association with breast cancer shown to be the likely result of bias.[61] A more recent report from an observational study has suggested a moderately high risk of cancer among hypertensive patients treated with a calcium antagonist.[62] But, once again, only limited data are currently available from randomised trials to assess the reliability of this observation.

Use of observational studies to estimate potential effects of treatment

Prediction of the relative effects of treatment

When a treatment alters an established risk factor for disease, observational studies of the association between that risk factor and the disease may provide some indication of the potential effects of the treatment on disease risk. For example, in observational studies, a prolonged 5 mm Hg lower diastolic blood pressure is associated with about a one-third lower risk of stroke among middle-aged individuals,[63] and randomised trials of blood-pressure lowering[26] show that much, or all, of this predicted long-term effect is achieved within 5 years (with similar relative treatment effects in a variety of subgroups of patients; figure). However, although such estimates from observational studies of the potential effects of treatments may be valuable, they could overestimate the actual effects of treatment if disease risk is only partly reversed (at least in the short term). For example, observational studies have shown that a prolonged 5 mm Hg lower diastolic blood pressure is associated with about a one-fifth lower risk of coronary heart disease,[63] and that a prolonged 1 mmol/L lower blood cholesterol concentration is associated with about a one-half lower risk of coronary heart disease[64] among middle-aged individuals. By contrast, randomised trials of treatments that reduce blood pressure[26] or cholesterol[64,65] suggest that only about a half or two-thirds of the long-term effects predicted from the observational studies are produced within about 5 years of altering these risk factors.

Treatments might also have independent effects on disease risk that offset or augment the benefits of altering a particular risk factor. For example, the reduction in stroke risk of about a third in a recent randomised trial of an angiotensin-converting-enzyme inhibitor[15] is about twice as great as would be predicted from observational studies[63] for the achieved 3 mm Hg reduction in systolic blood pressure. Moreover, differences in outcome associated with some putative risk factors may not be causal. For example, many observational studies have found that greater consumption, and higher blood concentrations, of β-carotene are associated with lower incidence of cancer. The first reviews of these findings stressed that such associations might merely reflect some type of confounding, and emphasised the need for large-scale randomised trials of the effects of long-term β-carotene supplementation on cancer incidence.[66,67] Even after more than 10 years of treatment in such trials, no clear evidence of benefit has emerged,[68] suggesting that the inverse associations in observational studies were



Meta-analysis of randomised controlled trials of antihypertensive therapy:[26] achievement of full effects on stroke risk predicted from observational studies[63] for a 5–6 mm Hg reduction in usual diastolic blood pressure

Diamonds=point estimates and 95% CIs. Arrows indicate range of effect predicted from observational studies. DBP=diastolic blood pressure.

indeed due largely, or wholly, to differences in other aspects of health-related behaviour, which determined differences in cancer risk between those with different β-carotene intakes. Similarly, blood concentrations of fibrinogen,[69] C-reactive protein,[69] and homocysteine[70] seem to be strongly associated with the risks of vascular disease, but it remains unclear whether blood concentrations of these factors are raised largely as a consequence of underlying vascular disease (rather than being a cause of it)—in which case, lowering the concentrations of these factors would not be expected to produce any material reduction in risk.

Prediction of the absolute effects of treatment

Another way in which observational studies may help to determine the potential effects of treatment is by providing more representative estimates of the absolute rates of death, and of other relevant outcomes, in particular circumstances in the absence of the treatment. For, although randomised controlled trials will usually provide the most reliable estimates of the relative effects of treatment on cause-specific outcomes,[1] restrictive inclusion criteria can result in the recruitment of patients at higher or lower than usual risk. In circumstances where the relative effects of treatment are similar across a wide range of disease risks (as, for example, with blood-pressure lowering and stroke risk; figure), the absolute effects of treatment will vary in approximate proportion to the background disease risks. So, for example, the likely absolute effects of antihypertensive treatment might be best estimated by applying the relative reductions in stroke and in coronary heart disease shown by the randomised trials[26] to the absolute rates of the same outcomes found in observational studies of specific populations. Hence, the absolute benefits of such treatment at the same levels of blood pressure would be expected to be much greater in populations with stroke and coronary disease rates that are very high (as in parts of eastern Europe and Russia,[71] or among patients with pre-existing vascular disease) than in those with much lower rates (as in parts of southern Europe,[71] or among patients without vascular disease or other important risk factors).

CONCLUSIONS: Improving health care by the appropriate use of epidemiological evidence

Both randomised trials and observational studies can contribute important evidence about the effects of treatment on mortality and major non-fatal outcomes. The appropriate role for each type of study is determined primarily by the extent to which random error and bias can be guaranteed to be sufficiently small for the question posed to be answered reliably. Observational studies can often reduce random error substantially by involving very large numbers of individuals with a specific disease outcome, thereby providing useful evidence about any large effects of treatment on relatively uncommon outcomes (for example, rare but serious side-effects). Such studies may also provide an indication of the eventual effects of a treatment that markedly alters levels of a risk factor, provided there is a causal relation with disease. But, due to their inherent potential for moderate or large biases, observational studies have little role in the direct assessment of any more moderate effects of treatment on major outcomes, which are generally all that can realistically be expected from most treatments for most common serious conditions. By contrast, random allocation of treatment minimises bias and, when random error is also reduced sufficiently by the study of appropriately large numbers of patients (whether in individual trials or meta-analyses), randomised trials can provide reliable evidence about moderate treatment effects.[1]

Failure to recognise the limitations of observational studies in the assessment of moderate treatment effects may well have serious consequences, including both the use of ineffective treatments (such as β-carotene supplements for cancer prevention[68]) and the inappropriate abandonment, or insufficiently widespread use, of effective treatments (as occurred, for example, when concerns were raised about the safety of certain antihypertensive drugs[72]). Despite this, it has been argued by some that observational studies can provide useful information when there are substantial barriers to the conduct of randomised trials, such as the requirement for an extremely large sample size or a very long period of follow-up (eg, assessment of the effects of hormone replacement therapy on breast-cancer risks); when the conduct of randomised trials is hindered by the reluctance of patients or their doctors to participate (eg, assessment of treatments for multiple sclerosis); or when there are considered to be other ethical, economic, regulatory, or political obstacles.[57,58] Difficulties in obtaining reliable evidence in randomised trials as a consequence of such obstacles are not, however, sufficient to justify the use of unreliable evidence from observational studies that may, due to the potential biases, be importantly misleading. Instead, greater efforts need to be made to remove or overcome any obstacles that inappropriately prevent the provision of reliable evidence from randomised trials of adequate size (as has been achieved for treatments of numerous vascular and neoplastic conditions).

It has also been argued that observational studies could provide more generalisable evidence about the effects of treatment because they involve populations of patients, or clinicians, that are more representative of particular practice settings than those involved in clinical trials.[57,58] But, the inclusion of more representative participants does not prevent observational studies from producing biased estimates of any moderate treatment effects that might exist. Moreover, as has been discussed, careful consideration of the effects seen among the different types of patient included in randomised trials can often allow the results of clinical trials to be generalised widely. In particular, applying estimates derived from appropriately large randomised trials (or meta-analyses of trials) of the relative effects of a treatment on specific outcomes to the absolute risks of those outcomes observed in representative patient populations may well provide a broadly reliable guide to the balance of the absolute benefits and absolute risks conferred by the treatment in routine clinical practice.

In conclusion, observational studies and randomised trials provide complementary evidence about the effects of treatment on mortality and major morbidity. Wider appreciation of the different strengths and weaknesses of these two types of epidemiological study should increase the likelihood that the most reliable evidence available informs decisions about the treatments doctors use—and patients receive—for the management of a wide range of life-threatening conditions.

We thank Colin Baigent, Iain Chalmers, Richard Doll, Richard Peto, Anthony Rodgers, and Peter Sandercock for comments and suggestions on previous drafts of this review. This work was supported by an Eccles Award from the National Health and Medical Research Council of Australia, a program grant from The Medical Foundation of the University of Sydney, a British Heart Foundation personal chair, and direct support from the UK Medical Research Council and Imperial Cancer Research Fund.

References

1 Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet* 2001; **357:** 373–80.

2 Mellin GW, Katzenstein M. The saga of thalidomide. *N Engl J Med* 1962; **267:** 1184–89.
3 Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med* 1971; **284:** 878–81.
4 Khan MA, Herzog CA, St Peter JV, et al. The prevalence of cardiac valvular insufficiency assessed by transthoracic echocardiography in obese patients treated with appetite-suppressant drugs. *N Engl J Med* 1998; **339:** 713–18.
5 Rzany B, Correia O, Kelly JP, Naldi L, Auquier A, Stern R, for the Study Group of the International Case-Control Study on Severe Cutaneous Adverse Reactions. Risk of Stevens-Johnson syndrome and toxic epidermal necrolysis during first weeks of antiepileptic therapy: a case-control study. *Lancet* 1999; **353:** 2190–94.
6 Grady D, Rubin SM, Petitti DB, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med* 1992; **117:** 1016–37.
7 Beral V, Banks E, Reeves G, Appleby P. Use of HRT and the subsequent risk of cancer. *J Epidemiol Biostats* 1999; **4:** 191–210.
8 Florey M, Florey H. General and local administration of penicillin. *Lancet* 1943; **1:** 387–97.
9 Harington M, Kincaid-Smith P, McMichael J. Results of treatment in malignant hypertension: a seven-year experience in 94 cases. *BMJ* 1959; **2:** 969–80.
10 Mahalanabis D, Choudhuri AB, Bagchi NG, Bhattacharya AK, Simpson TW. Oral fluid therapy of cholera among Bangladesh refugees. *Johns Hopkins Med J* 1973; **132:** 197–205.
11 Hennekens CH, Buring JE. Observational evidence. *Ann N Y Acad Sci* 1993; **703:** 18–24.
12 Lever AF, Hole DJ, Gillis CR, et al. Do inhibitors of angiotensin-I-converting enzyme protect against risk of cancer? *Lancet* 1998; **352:** 179–84.
13 Pfeffer MA, Braunwald E, Moyé LA, et al, on behalf of the SAVE investigators. Effect of captopril on mortality and morbidity in patients with left ventricular dysfunction after myocardial infarction: results of the survival and ventricular enlargement trial. *N Engl J Med* 1992; **327:** 669–77.
14 The SOLVD Investigators. Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. *N Engl J Med* 1992; **327:** 685–91.
15 The Heart Outcomes Prevention Evaluation Study Investigators. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. *N Engl J Med* 2000; **342:** 145–53.
16 Barrett-Connor E, Grady D. Hormone replacement therapy, heart disease, and other considerations. *Annu Rev Public Health* 1998; **19:** 55–72.
17 Barrett-Connor E. Hormone replacement therapy. *BMJ* 1998; **317:** 457–61.
18 Jick H, Zornberg GL, Jick SS, Seshadri S, Drachman DA. Statins and the risk of dementia. *Lancet* 2000; **356:** 1627–31.
19 Soumerai SB, McLaughlin TJ, Spiegelman D, Hertzmark E, Thibault G, Goldman L. Adverse outcomes of underuse of β-blockers in elderly survivors of acute myocardial infarction. *JAMA* 1997; **277:** 115–21.
20 Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985; **27:** 335–71.
21 Doughty RN, Rodgers A, Sharpe N, MacMahon S. Effects of beta-blocker therapy on mortality in patients with heart failure: a systematic overview of randomized controlled trials. *Eur Heart J* 1997: **18:** 560–65.
22 CIBIS-II Investigators and Committees. The Cardiac Insufficiency Bisoprolol Study II (CIBIS II): a randomised trial. *Lancet* 1999; **353:** 9–13.
23 Hjalmarson A, Goldstein S, Fagerberg B, et al, for the MERIT-HF Study Group. Effects of controlled-release metoprolol on total mortality, hospitalizations, and well-being in patients with heart failure: the Metroprolol CR/XL Randomized Intervention Trial in Congestive Heart Failure (MERIT-HF). *JAMA* 2000; **283:** 1295–302.
24 Clarke R, Shipley M, Lewington S, et al. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *Am J Epidemiol* 1999; **150:** 341–53.
25 Thurmer HL, Lund-Larsen PG, Tverdal A. Is blood pressure treatment as effective in a population setting as in controlled trials? Results from a prospective study. *J Hypertens* 1994; **12:** 481–90.
26 Collins R, MacMahon S, Blood pressure, antihypertensive drug treatment and the risks of stroke and of coronary heart disease. *Br Med Bull* 1994; **50:** 272–98.
27 Paganini-Hill A, Chao A, Ross RK, Henderson BE. Aspirin use and chronic diseases: a cohort study of the elderly. *BMJ* 1989; **299:** 1247–50.
28 Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy. I: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ* 1994; **308:** 81–106.
29 Barnett HJM, Taylor DW, Eliasziw M, et al, for the North American Symptomatic Carotid Endarterectomy Trial Collaborators. Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. *N Engl J Med* 1998; **339:** 1415–25.
30 Taylor DW, Barnett HJM, Haynes RW, et al, for the ASA and Carotid Endarterectomy (ACE) Trial Collaborators. Low-dose and high-dose acetylsalicylic acid for patients undergoing carotid endarterectomy: a randomised controlled trial. *Lancet* 1999; **353:** 2179–84.
31 Swan SH, Shaw GM, Schulman J. Reporting and selection bias in case-control studies of congenital malformations. *Epidemiology* 1992; **3:** 356–63.
32 Stewart A, Webb J, Hewitt D. A survey of childhood malignancies. *BMJ* 1958; **1:** 1495–508.
33 MacMahon B, Hutchinson GB. Prenatal x-ray and childhood cancer: a review. *Acta Unio Int Contra Cancrum* 1964; **20:** 1172–74.
34 Khoury MJ, James LM, Erickson JD. On the use of affected controls to address recall bias in case-controlled studies of birth defects. *Teratology* 1994; **49:** 273–81.
35 Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet* 1996; **347:** 1713–27.
36 Bar-Oz B, Moretti ME, Mareels G, Van Tittelboom T, Koren G. Reporting bias in retrospective ascertainment of drug-induced embryopathy. *Lancet* 1999; **354:** 1700–01.
37 Psaty BM, Heckbert SR, Koepsell TD, et al. The risk of myocardial infarction associated with antihypertensive drug therapies. *JAMA* 1995; **274:** 620–25.
38 Blood Pressure Lowering Treatment Trialists' Collaboration. Effects of ACE inhibitors, calcium antagonists, and other blood-pressure-lowering drugs: results of prospectively designed overviews of randomised trials. *Lancet* 2000; **355:** 1955–64.
39 McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 1994; **272:** 859–66.
40 Gargiullo PM, Kramarz P, DeStefano F, Chen RT. Principles of epidemiological research on drug effects. *Lancet* 1999; **353:** 501.
41 MacDonald TM, Evans JMM, Sullivan F, McMahon AD. Principles of epidemiological research on drug effects. *Lancet* 1999; **353:** 501–02.
42 Barbone F, McMahon AD, Davey PG, et al. Association of road-traffic accidents with benzodiazepine use. *Lancet* 1998; **352:** 1331–36.
43 Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; **127:** 757–63.
44 Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52 705 women with breast cancer and 108 411 women without breast cancer. *Lancet* 1997; **350:** 1047–59.
45 Matthews KA, Kuller LH, Wing RR, Meilahn EN, Plantinga P. Prior to use of estrogen replacement therapy, are users healthier than nonusers? *Am J Epidemiol* 1996; **143:** 971–78.
46 Barrett-Connor E. Postmenopausal estrogen and prevention bias. *Ann Intern Med* 1991; **115:** 455–56.
47 Giovannucci E, Colditz GA, Stampfer MJ, Willett WC. Physical activity, obesity, and risk of colorectal adenoma in women (United States). *Cancer Causes Control* 1996; **7:** 253–63.
48 Hulley S, Grady D, Bush T, et al, for the Heart and Estrogen/progestin Replacement Study (HERS) Research Group. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 1998; **280:** 605–13.
49 Larkin M. Ups and downs for HRT and heart disease. *Lancet* 2000; **355:** 1338.
50 Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; **342:** 1878–86.
51 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; **342:** 1887–92.
52 Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 2000; **342:** 1907–09.
53 Moss AJ, Davis HT, Conrad DL, DeCamilla JJ, Odoroff CL. Digitalis-associated cardiac mortality after myocardial infarction. *Circulation* 1981; **64:** 1150–56.
54 The Digitalis Investigation Group. The effect of digoxin on mortality and morbidity in patients with heart failure. *N Engl J Med* 1997; **336:** 525–33.
55 Lindsay SJ, Kearney MT, Prescott RJ, Fox KAA, Nolan J, for the UK Heart Investigation. Digoxin and mortality in chronic heart failure. *Lancet* 1999; **354:** 1003.
56 Spargias KS, Hall AS, Ball SG. Safety concerns about digoxin after acute myocardial infarction. *Lancet* 1999; **354:** 391–92.

57 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; **312:** 1215–18.

58 McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999; **319:** 312–15.

59 Report from the Boston Collaborative Drug Surveillance Program. Reserpine and breast cancer. *Lancet* 1974; **2:** 669–71.

60 Heinonen OP, Shapiro S, Tuominen L, Turunen MI. Reserpine use in relation to breast cancer. *Lancet* 1974; **2:** 675–77.

61 Horwitz RI, Feinstein AR. Exclusion bias and the false relationship of reserpine and breast cancer. *Arch Intern Med* 1985; **145:** 1873–75.

62 Pahor M, Guralnik JM, Ferrucci L, et al. Calcium-channel blockade and incidence of cancer in aged populations. *Lancet* 1996; **348:** 493–97.

63 MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990; **335:** 765–74.

64 Law MR, Wald NJ, Thompson SG. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? *BMJ* 1994; **308:** 367–73.

65 LaRosa JC, He JM, Vupputuri S. Effect of statins on risk of coronary disease: a meta-analysis of randomized controlled trials. *JAMA* 1999; **282:** 2340–46.

66 Peto R, Doll R, Buckley JD, Sporn MB. Can dietary beta-carotene materially reduce human cancer rates? *Nature* 1981; **290:** 201–08.

67 Buring J, Hennekens C. Retinoids and carotenoids. In: De Vita VT Jr, Hellman S, Rosenberg S, eds. Cancer: principles and practice of oncology. Philadelphia: J B Lippincott, 1993.

68 Hennekens CH, Buring JE, Manson JE, et al. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N Engl J Med* 1996; **334:** 1145–49.

69 Danesh J, Collins R, Appleby P, Peto R. Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies. *JAMA* 1998; **279:** 1477–82.

70 Danesh J, Lewington S. Plasma homocysteine and coronary heart disease: systematic review of published epidemiological studies. *J Cardiol Risk* 1998; **5:** 229–32.

71 WHO. World health statistics annual. Geneva: WHO, 1996.

72 Ad Hoc Subcommittee of the Liaison Committee of the World Health Organisation and the International Society of Hypertension. Effects of calcium antagonists on the risks of coronary heart disease, cancer and bleeding. *J Hypertens* 1997; **15:** 105–15

The uses of error: The complexity of general practice

General practitioners encounter a large number and variety of health problems every day. In coping with these problems the general practitioner has to modify his or her role accordingly, from masterly inactivity to active diagnostic and therapeutic pursuit. This case is an example of complex role-modification leading to a tragic, though probably understandable, failure to diagnose breast cancer.

A 37-year-old patient used to visit my practice regularly for various minor problems. I was rather fond of her and admired her for her cheerful outlook of life: she was a young mother of a single teenage son, and it was no secret that she was less than happy with her elderly husband. She had consulted me for a lump in her right breast, which appeared benign on initial examination.

This was the planned follow-up 3 weeks later to check on its natural history.

"How about the breast?" I started. "Yes, well, the breast," she replied, her body language signalling utter confusion. This rang alarm bells inside my head so I invited her to tell me what was troubling her. She explained her recent discovery of being lesbian, the delight of finding a girlfriend, and her anxiety regarding the consequences this had for her own family. I was allowed to check her breast, and I happily concluded that the lump had reduced in size, if not disappeared. I urged her to come back within a month. However, during that consultation as well as the next, we discussed her new perspective on life and the impossibility of integrating it with her previous one. It was only after a number of such highly emotional consultations that I dared to request permission to re-examine her breast. The examination disclosed a persistent lump in the right breast, and I managed to arrange for a consultation with a surgeon next day. The outcome was as striking as it was predictable. A carcinoma was diagnosed, an operation performed, 8 months later metastases were discovered, and just over a year later she died.

Two memories of that period prevail. It was a considerable shock to receive a warm welcome when I visited her in hospital after the operation, because I had been expecting remorseless reproach of my delayed diagnosis. The other memory is of the complicated, yet highly rewarding, period of terminal home care including a still fundamentally puzzled ex-husband and a most supportive girlfriend.

This case exemplifies the complexity of general practice. A lump in the breast is a common presenting complaint and important life events are often discussed between patients and their family doctor. This example highlights the difficulty of simultaneously pursuing a clinical diagnosis while providing a sympathetic ear for a patient's often unrelated difficulties, a duty that the medical profession must carry out on both accounts.

Professor Chris van Weel

Department of General Practice, University of Nijmegen, Netherlands