

Design issues for drug epidemiology

Alex D. McMahon¹ & Tom M. MacDonald²

¹Robertson Centre for Biostatistics, University of Glasgow, Glasgow G12 8QQ and ²Medicines Monitoring Unit, University of Dundee, Dundee DD1 9SY

Despite the difficulties involved in designing drug epidemiology studies, these studies are invaluable for investigating the unexpected adverse effects of drugs. The aim of this paper is to discuss various aspects of study design, particularly those issues that are not easily found in either textbooks or review papers. We have also compared and contrasted drug epidemiology with the randomized controlled trial (RCT) wherever possible. Drug epidemiology is especially useful in the many situations where the RCT is not suitable, or even possible. The study base has to be defined before the appropriate cohort of subjects is assembled. If all of the cases are identified, then a referent sample of controls may be assembled by random sampling of the study base. If all of the cases cannot be assembled, a hypothetical secondary base may need to be created. Preferably, only new-users of the drug should be included, and the risk-ratio will be different for acute users and chronic users. Studies will usually only be possible when researching the unintended effects of drugs. It is difficult to study efficacy because of confounding by indication. In occasional circumstances it may be possible to study efficacy (examples are given). Discussion of the dangers of designing with generalisability in mind is provided. Additionally, the similarities in study design between drug epidemiology and the RCT are discussed in detail, as well as the design-characteristics that cannot be shared between the two methods.

Keywords: drug epidemiology, pharmacoepidemiology, study design.

Introduction

Designing studies in drug epidemiology can be a very complicated business. Despite the difficulties, many useful studies have been performed that combined pharmaco-epidemiological techniques with large record-linkage databases of drug prescriptions and clinical outcomes [1, 2]. Generally speaking, there is a consensus amongst modern epidemiologists with regard to epidemiological methods, and study design in particular [3]. Even though there are recent controversies raging in the world of epidemiology [4], more basic errors such as the desire for controls in a case-control study to be 'very healthy' are still being made [5]. Against a backdrop of recent confusion in (specifically) drug epidemiology, a leading researcher decided it was necessary to publish guidelines for the design of such studies [6], even though something similar

had been published in 1978 [7]. The purpose of the current paper is to discuss various aspects of study design that are of interest to the authors. We have attempted to highlight issues that are not easily found in either textbooks or review papers. We have also compared and contrasted drug epidemiology with the randomized controlled trial wherever possible.

Why use drug epidemiology?

Although the randomized controlled trial is the best way to demonstrate the effects of pharmaceuticals, these trials are expensive to carry out. It is probable that the majority of medical research is carried out using purely observational data, i.e. no trial intervention or control was used. Some researchers have even criticised the dominance of the controlled trial, and even randomization itself, as the best methodology for carrying out clinical research [8, 9]. It is sometimes forgotten that the real strength of randomization is not to be found in obscure statistical subtleties, it is the ability to demonstrate cause and effect [10]. Because of

Correspondence: Alex McMahon, Robertson Centre for Biostatistics, Boyd Orr Building, University of Glasgow, Glasgow, G12 8QQ, Scotland.

Received 5 April 2000, accepted 22 August 2000.

the many sources of bias in an observational study, demonstrating cause and effect is much more difficult and it is often necessary to rely on evidence from outwith a study. In a very old paper by Cochrane [11], an anecdote is provided in which the eminent statistician Sir Ronald Fisher was asked what could be done in observational studies to make that step to causation. Fisher's answer, in seeming contradiction of Occam's Razor (i.e. the simplest answer is the most likely), was to 'make your theories elaborate'. He meant that observational studies have to consider as many different reasons for an association that can be thought of, and try and rule out as many as possible.

Although the importance of randomization cannot be overestimated, there are many situations where the randomized controlled trial (RCT) is not suitable. This may be due to either the practicalities of carrying out a study, or because randomization is often unethical. For example, a clinical trial can only detect very frequent adverse drug effects, as too few subjects will typically be studied [12, 13] and the observational methods of pharmacovigilance (as invented by Finney [14]) have to be used instead [15]. Other examples of difficulties with RCTs include the following [16]: RCTs are not useful for studying drug interactions or genetic disposition to diseases; drugs are used for indications other than the licensed ones; it may be unethical to randomize some groups of people such as pregnant women or children (although this principal may be misused when the real reason for exclusion is merely convenience); and finally, it is not ethical to design an RCT that examines drug overdose.

Defining the cohort

The first thing to do when a hypothesis has been developed, and study design has begun, is to define the cohort within which the study will be conducted. The *source population* is the population which has the available data and measurements that are required to answer the questions. The *study population* is the remaining population after all inclusion and exclusion criteria have been applied. When the period of time over which the study population will be studied is added to the equation, we then have the *study base*. The study base is therefore the members of the study cohort during the time that will be used in the study [17–19]. A cohort study is constructed explicitly within the study base by examining exposure patterns and looking for an association with the disease (or 'event', or 'outcome') of interest. Defining the base is also important for case-control studies, because these studies also take place within the study base. When described this way, it is obvious that a case-control study is merely a special type of sample of the underlying cohort [20]. If a case-control

study is to be carried out then a complete census of all of the events, or *cases*, must be assimilated into the case-series. At the very least, the case-series should be a 'random' sample of all of the cases [17, 21]. In other words, the cases in a case-control study are the same cases that would be used in an equivalent cohort study. The difference between a cohort study and a case-control study is that in the latter type of study we *sample* the cohort to provide a set of controls [5].

The secondary base

Sometimes, trying to imagine the study base for a set of cases that have already been assembled can be difficult, although this exercise is instructive in itself [22]. Alternatively, the base may be defined but we cannot identify all of the cases, and we may only be able to identify a particular *subset* of the cases. Either of these two scenarios fails the simple rules above. A valid study might still be possible when this happens by imagining a *secondary base*, although the study might not be representative of a real population when doing this. The secondary base is an artificial base for which the set of cases is indeed complete [17, 21].

The most common example of this is the hospital-based case-control study, where the available cases are only those cases which 'ended up' in a particular hospital [18, 23]. The controls may have to be drawn from either the hospital catchment area, so that they would have gone to hospital if they developed the disease, or they must be actual patients in the hospital, if the patients in the catchment area would not necessarily have been in the case-series if they developed the disease. This type of study is easier to understand when data are available for *all* of the hospitals in a particular population. At least there is no concern over the case-mix of a particular hospital; we only worry about the possibility of cases occurring in the community not reaching hospital at all. For serious diseases, it may be possible to argue that the vast majority of cases have been identified (with the possible addition of community death-registration data).

Note that if hospitalized controls are being used, it is important to exclude patients who have been admitted for a disease that is associated with the study drug in order to prevent selection bias. Also, in all studies the subjects with prior study-events will usually have to be excluded because of contraindication with the study drug [7, 24], unless the effect of the drug (usually an adverse effect) is hitherto unknown [25]. Generally, subjects who have evidence of illnesses that are risk factors for the outcome of interest, and for which the study drug is either indicated for, or indicated against (contraindicated), should also be excluded [6]. An example would be a study of the association between nonsteroidal anti-inflammatory drugs

(NSAIDs) and gastric bleeding. Previous endoscopic examination will be a risk factor for bleeding, and will also be a contraindication for prescribing of NSAIDs. Therefore, including these patients would bias the size of the toxic effect downwards, so patients with prior endoscopies should be excluded.

It may be tempting to keep contraindicated subjects in a study when the researcher is struggling with an under-powered sample-size. There could be a desire to settle a serious drug-toxicity problem as early as possible. Although the study may indeed have more power by including a larger number of events, doing this will create other statistical difficulties. Risk factors that are contraindications will produce very strong interaction effects. Essentially, the increased rate of adverse events that is due to exposure, will only manifest itself in the subjects *without* contraindications, although the event rate may be higher in these subjects. For example, in a previous study of the gastric toxicity of NSAIDs [26], the incidence rate (per thousand person years of exposure) for subjects with prior ulcer healing drugs was 8.02 for exposed subjects, and 9.08 for unexposed subjects. For subjects *without* prior ulcer healing drugs, the incidence rate was 4.45 for exposed subjects and 1.61 for unexposed subjects. Therefore the rate-ratios for exposure *vs* nonexposure were 0.88 for subjects with this particular contraindication to NSAIDs, and 2.77 for subjects without the contraindication. This is a very strong interaction ($P=0.004$), which means that we shouldn't combine these subgroups.

Some of the published reasons for exclusions in a study are not essential for the purposes of removing dangerous sources of bias. It has been suggested that we should only include cases of uncertain cause [6], and exclude cases with an 'alternate proximate cause'. This is a bit like censoring the irrelevant deaths in a survival analysis (for example in a Kaplan-Meier plot). These alternate cases may not necessarily be biased towards either exposed subjects or unexposed subjects. When this type of exclusion is considered the analysis may be carried out both ways, including and excluding these cases, out of curiosity about what happens when this possible source of error (i.e. 'noise') has been removed from the results.

New users

It is important to consider the timing of events in relation to the start of drug exposure [27]. Another criterion that should be applied to drug epidemiology studies of drug-toxicity is that a study should only include new users of the drug of interest. Any previous adverse experiences with a drug will be a contraindication to future exposure [24]. Similarly, past use of a drug, and chronic prescribing in particular, will tend to be associated with nonsusceptibility to any adverse effect of the drug [28]. This means that we

should expect the risk of an event to be higher in acute users of a drug than in more chronic users [24, 27–29]. It could be argued that, in some studies, the previous use of a drug might cause a relatively *mild* form of confounding by contraindication. Some researchers may carry out at least one analysis that includes the previous users due to concerns with the study being under-powered. If the study was one of drug-efficacy (assuming that a study was possible) the role of previous prescribing would not be as obvious. There could be a selection bias due to the differing effects of diagnoses in the 'distant-past' and more recent diagnoses, so perhaps a study of efficacy should be analysed separately for these different types of patient, as a check of whether the results are affected.

For the stated reasons, it is not desirable to mix together subjects with different patterns of drug usage [30]. Also, if acute users and chronic users are mixed together then the important assumption that the hazard rate is constant across time may be violated [27]. This assumption underlies both cohort and case-control studies, and can also be broken if evolving clinical practice creates change in the pattern of contraindications [31]. Analyses that examine the cumulative effects of repeated prescriptions should be restricted to only those subjects with chronic prescribing [29], and only after establishing the start of exposure. New-use of the drug can be established by creating an 'inception cohort' [32] from the date of the first prescription. In practice this might be difficult if a database begins collecting data from a particular date, in which case a sacrificial screening period may be used to screen out past users [26, 33].

Intended effects

The greatest successes of epidemiology have been with the unintended effects of exposures such as adverse effects [25], and much less has been learned about the intended effects, such as efficacy. In drug epidemiology this is due to powerful biases that confound an association with the indication for a treatment. If we wish to compare treated patients with untreated patients, and we assume that prescribing is rational, then the treated patients will automatically have a higher rate of any disease that the drug is intended to treat (or possibly cure) [34]. Therefore a drug that actually helps patients will appear to be risky.

This means that studying efficacy with observational data is extremely difficult, and usually impossible. This can be seen in a positive light; the reverse is true of randomized clinical trials, because they are an excellent method of examining efficacy, and are not usually suitable for looking at unintended effects. As an aside, this should serve as a warning to the emerging field of 'outcomes research', which examines the 'effectiveness' of health technologies [25, 35]. The usefulness of routinely collected medical

data, for the evaluation of treatments or medical interventions in general, is likely to be limited.

Despite the problems, there is clearly a demand for epidemiological studies of drug-effectiveness, especially for drugs that are being used for unlicensed indications [36]. There are in fact some situations where nonexperimental methods could be used to demonstrate efficacy [37]. The effect of the drug could be so dramatic that no comparator group is required, for example the use of naloxone in patients who are comatose with opiate poisoning [34]. A disease could be stable or predictable so that fluctuations may be attributable to an exposure; for example insulin use and glycaemic control in diabetes [38]. However, drugs can appear efficacious simply due to 'regression to the mean', when patients will get well over time, especially if selected into a cohort at a time of severe illness. In theory, if the severity of an indication could be measured exactly then this could be adjusted for in an analysis, although this is usually not possible [39, 40]. In practice, epidemiological studies of drug-efficacy suffer from uncontrollable bias due to confounding by indication.

Generalisability

When designing a study we have seen that various exclusions have to be made in order to create a study base that is as unbiased as possible. A large part of a source population may be thrown out of a study in this way [41]. This is paralleled by the inclusion and exclusion criteria used in clinical trials. The subjects in an RCT will not usually be a random sample of any population at all, although they are often considered to be just that. What is important is whether or not the treatment works for those patients who were randomized. The result is still valid using "proof within the trial" and the results only provide an approximation to what will happen in a true population [10, 42]. Emphasis on the 'representativeness' of the subjects in either a clinical trial [42] or an observational study can be damaging to the design of a study [17, 41, 43]. It should be safer to generalize a treatment difference (e.g. active *vs* placebo) than the actual success rate in the treated subjects [44].

There are some aspects of study design with clinical trials that are worth trying to emulate in an observational study. Let us consider the RCT again. As was mentioned, the primary purpose of a clinical trial is not to be representative of a population, it is to find a *difference* between treatments (equivalence studies are being ignored, because 'equivalence is different' [45]). This simple point has firm grounds in the philosophy of science [10]. Simply put, we are trying to refute the suggestion that the null hypothesis is true, and *not* trying to prove that the *alternative* is true. Having said that, it is desirable to have as wide a selection of patients in a trial as possible, as we are

only excluding subjects who cannot be randomized for ethical reasons. There have often been justified criticisms of RCTs, stating that they are usually excessively restrictive in their choice of subjects [44, 46–49]. In this respect the RCT does not follow the paradigm of experimental science because the experimental 'units' are not intended to be homogenous (e.g. genetically similar rats). Some researchers may disagree with this view, but we remain convinced.

The main criticism is that if RCTs exclude certain types of patient, then this makes it difficult to extrapolate the results. As we have said, although patients should not be unnecessarily excluded, this external application of study-results may often be less problematic than doctors may think. Simply put, the treatment effect may be smaller or larger for patients with different prognostic factors, rather than nonexistent or even harmful for some patients [46, 50]. Ironically, when a study *does* have wide entry criteria, some researchers may attempt to claim that the drug does not work for a particular subgroup! This is of course an unacceptable practice, and one study has humorously pointed out that their drug did not appear to work for patients born under two of the astrological birth signs [51]. The most likely treatment response in a subgroup is that for the whole study [50].

Where an RCT does follow the experimental paradigm is in the idea of experimental control. In an RCT, the conduct of the study has to be tightly controlled in order to give the best chance of *detecting a difference between the treatments*. This is why it is more difficult to find a treatment effect with either an intention-to-treat analysis, because the study conduct was not as good as that hoped for, or in a so-called 'pragmatic' study. In an observational study exclusions may be made in the spirit of a clinical trial [52, 53], but for very different reasons. Because there is no randomization of treatments, the only alternative that may help in preventing bias is 'judicious selection of subjects' [54]. Restricting entry into the study in ways similar to the RCT can make the results less biased, and in some cases very like the results of equivalent RCTs [55]. Although the reason for exclusions is different, researchers should be wary of over-reliance on the flawed concept that study subjects should be representative of a population [7, 17, 41, 43] and not be afraid of 'throwing away data'. In the past authors have over-emphasized the virtues of generalisability with observational research [56]. As we have discussed, perhaps observational studies have not always been restrictive enough and RCTs have often been *too* restrictive, which raises the possibility that there is an optimum level of subject-restriction that is common to both methods.

Directionality

At one time case-control studies were ridiculed as inferior retrospective 'trohoc' studies (i.e. cohort spelt backwards) [57]. Nowadays it is recognized that both cohort and case-control studies may be conducted either prospectively or retrospectively. The idea that the concept of 'directionality' can distinguish between the two is now rejected as 'founded on nonsense' [58, 59]. In practice, most observational research is probably retrospective since this type of research often takes advantage of data that has been recorded for other purposes. However, thinking about directionality can help in designing a study, because observational research is usually theoretically 'retrospective' for an entirely different reason.

In an RCT the anchor of time upon which the study is built, is the point of randomization. An observational study does not have this anchor, and will usually be centred on the time of the outcome, at the *end* of the study. This may seem counter to intuition, but consider the following quote concerning cohort studies from over 20 years ago by the eminent epidemiologist Liddell; 'our research is prospective more in appearance than in fact, and as in all such studies, it is only possible, in practice, to classify the subjects' exposure in terms of length and intensity after it has ended' [60]. In other words, although a study may be conducted prospectively, and cause and effect reasoning is always in a forwards direction, observational variables will usually have to be measured backwards from the outcome [58].

The idea of a 'baseline' may not therefore mean much in drug epidemiology because the exposure might actually be in transit through the hypothetical baseline, which is really just a 'study start day' when data recording began. Exceptions might be studies where the baseline (i.e. 'time zero') could be created because the study also *begins* with a disease (e.g. myocardial infarction, and subsequent recurrence [55]). In theory a record-linkage system that covers an entire population, could simply stay in business until all of the subjects' prescriptions are recorded, from birth to death.

There is also no way to force exposure to be constant in the manner of a controlled trial (notwithstanding the usual problems of patient compliance). That exposure cannot be kept constant, and time is anchored at the outcome, can often be virtues of observational research. A parallel-group clinical trial could never discover that it is the blood alcohol immediately before a road accident that matters, or that it is the physical activity in the hours before a myocardial infarction that is the cause. The earlier exposure in these two examples did not matter at all.

The clinical trial paradigm

We have noted that, firstly, a clinical trial is anchored around the time of randomization and that exposure is held constant throughout the study and at the time of an outcome, and secondly, observational studies are usually assessed backwards from the outcome. Although there are many useful lessons from the clinical trial, these points mean that the clinical trial ultimately fails as a paradigm for drug epidemiology [41, 61]. This discovery certainly met some resistance when it was first suggested [62], and is probably still not appreciated by many researchers today.

In conclusion, the importance of randomization cannot be overestimated. In many situations a drug epidemiology study may be difficult, or even impossible to carry out. Indeed, anticipating when studies are impossible is an important skill in study design. However, it can also be difficult or impossible (at least ethically) to carry out a randomized controlled trial. Happily, these are often the very situations when drug epidemiology is at its most useful.

References

- 1 Evans JMM, MacDonald TM. Record-linkage for pharmacovigilance in Scotland. *Br J Clin Pharmacol* 1999; **47**: 105-110.
- 2 Garcia Rodriguez LA, Guthann SP. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol* 1998; **45**: 419-425.
- 3 Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd edn. Lippincott-Raven, Philadelphia, 1998.
- 4 Gori GB. Epidemiology and public health: is a new paradigm needed or a new ethic? *J Clin Epidemiol* 1998; **51**: 637-641.
- 5 Poole C. Controls who experienced hypothetical causal intermediates should not be excluded from case-control studies. *Am J Epidemiol* 1999; **150**: 547-551.
- 6 Jick H, Garcia Rodriguez LA, Perez-Guthann S. Principles of epidemiologic research on adverse and beneficial drug effects. *Lancet* 1998; **352**: 1767-1770.
- 7 Jick H, Vessey MP. Case-control studies in the evaluation of drug-induced illness. *Am J Epidemiol* 1978; **107**: 1-7.
- 8 Herman J. Experiment and observation. *Lancet* 1994; **344**: 1209-1211.
- 9 Vandenbrouke JP. Is the randomised controlled trial the real paradigm in epidemiology? *J Chronic Dis* 1986; **39**: 572.
- 10 Senn SJ. Falsification and clinical trials. *Statistics Med* 1991; **10**: 1679-1692.
- 11 Cochran WG. The planning of observational studies of human populations (with discussion). *J Royal Statist Soc, Series A* 1965; **128**: 234-265.
- 12 Carson JL, Strom BL, Maislin G. Screening for unknown effects of newly marketed drugs. In *Pharmacoepidemiology* (2nd edn). Chapter 30, eds. Strom BL. 1994.
- 13 Waller PC, Wood SM, Breckenbridge AM, Rawlins MD. Why the safety assessment of marketed medicines (SAMM) guidelines are needed. *Br J Clin Pharmacol* 1994; **38**: 93.

- 14 Finney DJ. The design and logic of a monitor of drug use. *J Chronic Dis* 1965; **18**: 77–98.
- 15 Lawson DH. Pharmacovigilance in the 1990s. *Br J Clin Pharmacol* 1997; **44**: 109–110.
- 16 Spilker. Chapter 3: 36–37. In *Pharmacoepidemiology an Introduction* eds. Hartzema AG, Porta MS, Tilson HH. Harvey Whitney Books, 1991.
- 17 Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I Principles. *Am J Epidemiol* 1992; **135**: 1019–1028.
- 18 Miettinen OS. The 'case-control' study: valid selection of subjects. *J Chronic Dis* 1985; **38**: 543–548 (part of a discussion section in, pp 541–558).
- 19 Miettinen OS. Cohorts versus dynamic populations: a dissenting view (Response to Greenland). *J Chronic Dis* 1986; **39**: 567.
- 20 Maclure M. Taxonomic axes of epidemiologic study designs: a refutationist perspective. *J Clin Epidemiol* 1991; **44**: 1045–1053.
- 21 Miettinen OS. The concept of secondary base. *J Clin Epidemiol* 1990; **43**: 1017–1020.
- 22 Knottnerus JA. Subject selection in hospital-based case-control studies. *J Chronic Dis* 1987; **40**: 183–185.
- 23 Miettinen OS. Subject selection in case-referent studies with a secondary base. *J Chronic Dis* 1987; **40**: 186–197.
- 24 Miettinen OS, Caro JJ. Principles of nonexperimental assessment of excess risk, with special reference to adverse drug reactions. *J Clin Epidemiol* 1989; **42**: 325–331.
- 25 Miettinen OS. The need for randomization in the study of intended effects. *Statistics Med* 1983; **2**: 267–271.
- 26 McMahon AD, Evans JM, White G, *et al.* A cohort study (with re-sampled comparator groups) to measure the association between new NSAID prescribing and upper gastrointestinal haemorrhage and perforation. *J Clin Epidemiol* 1997; **50**: 351–356.
- 27 Guess HA. Behaviour of the exposure odds ratio in a case-control study when the hazard function is not constant over time. *J Clin Epidemiol* 1989; **42**: 1179–1184.
- 28 Moride Y, Abenheim L. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol* 1994; **47**: 731–737.
- 29 McMahon AD, Evans JMM, McGilchrist MM, McDevitt DG, MacDonald TM. Drug exposure risk windows and unexposed comparator groups for cohort studies in pharmacoepidemiology. *Pharmacoepidemiol Drug Safety* 1998; **7**: 275–280.
- 30 Gerstman BB, Lundin FE, Stadel BV, Faich GA. A method of pharmacoepidemiologic analysis that uses computerized MEDICAID. *J Clin Epidemiol* 1990; **43**: 1387–1393.
- 31 Joseph KS. The evolution of clinical practice and time trends in drug effects. *J Clin Epidemiol* 1994; **47**: 593–598.
- 32 Hartzema AG. Guide to interpreting and evaluating the pharmacoepidemiologic literature. *Ann Pharmacother* 1992; **26**: 96–98.
- 33 Mantel N. Avoidance of bias in cohort studies. *Natl Cancer Inst Monograph* 1985; **67**: 169–173.
- 34 Strom B, Miettinen OS, Melmon KL. Postmarketing studies of drug efficacy: when must they be randomised? *Clin Pharmacol Ther* 1983; **34**: 1–7.
- 35 Davies HTO, Crombie IK. Outcomes from observational studies: understanding causal ambiguity. *Drug Information J* 1999; **33**: 151–158.
- 36 Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: why? *Am J Med* 1985; **78**: 475–480.
- 37 Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? *Am J Med* 1984; **77**: 703–708.
- 38 Morris AD, Boyle DIR, McMahon AD, Greene SA, MacDonald TM, Newton RW. Adherence to insulin treatment, glycaemic control, and ketoacidosis in insulin-dependent diabetes mellitus. *Lancet* 1997; **350**: 1505–1510.
- 39 Shapiro S. Case-control surveillance. In *Pharmacoepidemiology* (2nd edn), ed Strom BL. Chichester, Wiley 1994.
- 40 Byar DP. Problems with using observational databases to compare treatments. *Statistics Med* 1991; **10**: 663–666.
- 41 Miettinen OS. The clinical trial as a paradigm for epidemiologic research. *J Clin Epidemiol* 1989; **42**: 491–496.
- 42 Senn SJ. Clinical trials and epidemiology. *J Clin Epidemiol* 1990; **43**: 628–631.
- 43 Miettinen OS. Evidence in medicine: invited commentary. *Can Med Assoc J* 1998; **158**: 215–221.
- 44 Altman DG, Bland JM. Generalisation and extrapolation. *Br Med J* 1998; **317**: 409–410.
- 45 Senn S. *Statistical Issues in Drug Development*; Chichester, John Wiley, p. 211, 1997.
- 46 Yusuf S, Held P, Teo KK. Selection of patients for randomised controlled trials: implications of wide or narrow eligibility criteria. *Statistics Med* 1990; **9**: 73–86.
- 47 McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomised and non-randomised studies. *Br Med J* 1999; **319**: 312–315.
- 48 Fulks A, Weijer C, Freedman B, Shapiro S, Skrutkowska M, Riaz A. A study in contrasts: eligibility criteria in a twenty-year sample of NSABP and POG clinical trials. *J Clin Epidemiol* 1998; **51**: 69–79.
- 49 Black N. Why we need observational studies to evaluate the effectiveness of health care. *Br Med J* 1996; **312**: 1215–1218.
- 50 Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomised clinical trials. *JAMA* 1991; **266**: 93–98.
- 51 ISIS-2(Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin both or neither among 17 187 cases of suspected acute myocardial infarction. ISIS-2. *Lancet* 1988; **ii**: 349–360.
- 52 Feinstein AR, Horwitz RI. Double standards, scientific methods and epidemiologic research. *N Engl J Med* 1982; **307**: 1611–1617.
- 53 Mayes LC, Horwitz RI, Feinstein AR. A collection of 56 topics with contradictory results in case-control research. *Int J Epidemiol* 1988; **17**: 680–685.
- 54 Rothman KJ. Epidemiologic methods in clinical trials. *Cancer* 1977; **39**: 1771–1775.
- 55 Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990; **89**: 630–638.
- 56 Hlatky MA, Lee KL, Harrell FE, *et al.* Tying clinical research to patient care by use of an observational database. *Statistics Med* 1984; **3**: 375–384.
- 57 Feinstein AR. Clinical Biostatistics XX. The epidemiologic

- trohoc, the ablative risk ratio and 'retrospective' research. *Clin Pharmacol Ther* 1973; **14**: 291–307.
- 58 Miettinen OS. Striving to deconfound the fundamentals of epidemiologic study design. *J Clin Epidemiol* 1988; **41**: 709–713.
- 59 Greenland S, Morgenstern H. Classification schemes for epidemiologic research designs. *J Clin Epidemiol* 1988; **41**: 715–716.
- 60 Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis: appraisal by application to asbestos mining. *J Royal Statist Soc, Series A* 1977; **140**: 469–491.
- 61 Miettinen OS. Unlearned lessons from clinical trials: a duality of outlooks. *J Clin Epidemiol* 1989; **42**: 499–502.
- 62 Feinstein AR. Unlearned lessons from clinical trials: a duality of outlooks. *J Clin Epidemiol* 1989; **42**: 497–498.