Journal of Epidemiology and Community Health 1990; 44: 179–186

REVIEW ARTICLE: *Research methods in epidemiology, V*

Bias in case-control studies. A review

Jacek A Kopec, John M Esdaile

It has been widely accepted that one reason for inconsistent or contradictory results of epidemiologic studies is bias. Therefore, an appreciation of potential sources of bias has become a critical issue in epidemiology. A large number of different sources and possible mechanisms of bias have been described but a consistent terminology has not evolved. Different authors have used different terms when referring to essentially the same type of bias. For example, Kleinbaum *et al*[1] and Rothman[2] classify biases into *selection* bias, *information* bias and *confounding*. Sackett offered a more detailed classification, with five major types of *sampling* bias (eg, admission rate bias, membership bias) and four types of *measurement* bias (eg, diagnostic suspicion bias, recall bias).[3] Feinstein distinguishes, inter alia, among *susceptibility* bias, *performance* bias, *transfer* bias and *detection* bias.[4] Last's *Dictionary of epidemiology*[5] gives definitions of 26 biases but fails to mention many of the terms proposed by other authors.

It is not our objective to provide a comprehensive list of biases, or to propose a new, improved taxonomy. Our purpose is to review the most common threats to validity in the design of case-control studies—commonly regarded as being particularly susceptible to bias—and to discuss some useful strategies in dealing with these problems. Certain issues will not be addressed, notably methods of controlling confounding, and the statistical analysis of case-control studies.

Defining a case-control study

Traditionally, a case-control (case-referent) study has been defined as an investigation into a relationship between a given disease and one or more causal or preventive factors, in which persons selected because they have the disease (the cases) and suitably selected persons who do not have the disease (the controls) are compared in terms of their exposure to the factors under study.[5–7] Case-control studies are usually considered to be an alternative to cohort studies, from which they are distinguished by two features: (1) sampling by disease as opposed to sampling by exposure, and (2) investigative movement from effect to cause as opposed to from cause to effect.[1 6–8] Recently, Kramer and Boivin[9] have postulated that the order of measurement ("directionality"), rather than the method of sampling, provides an "unconfounded" criterion for distinguishing between the two designs. This approach is not shared by other authors. For Greenland and Morgenstern, case-control studies are distinguished from cohort studies by their use of "outcome-selective sampling".[10] Miettinen rejects the case-control versus cohort duality entirely. In his view, the "case-referent strategy" consists in: (1) identifying and classifying all of the cases in the "study base"[a], defined as the "population experience captured in a study", and (2) drawing a sample of the base to obtain information on the exposure.[12] A good illustration of this concept is a "nested", or "within cohort" case-control study, which can be regarded as an efficient form of the cohort design, rather than an entirely different type of study.[13]

It may be noted that the two aspects (or "axes") of the design, ie, order of measurement and method of sampling, are interdependent. The reason for determining exposure status for all subjects in most cohort studies is to obtain as valid information as possible for the relatively few cases-to-be. Therefore, in cohort studies all *potential* cases are classified as exposed or unexposed. The logic behind the case-control method is to make the design more efficient by classifying only the *actual* cases, since information about the source population can be obtained by sampling. This, however, implies a particular order of measurement (directionality), namely ascertaining the outcome first. Measuring exposure for all members of the study population after the cases have been ascertained is unnecessary and impractical, if the disease is rare.

The RCT paradigm

There seems to be current agreement that the methods of the randomised clinical trials may serve as a useful paradigm in designing case-control studies. Yet for many clinicians, the conceptual links between the randomised clinical trial model and the case-control design may not be obvious. As an illustration, consider the famous study of cholera in London, conducted by John Snow in the mid-nineteenth century.[14] The study has often been referred to as a "natural experiment",[2] and was practically as valid as a randomised controlled trial. However, its design resembled a modern, nested case-control ("case-cohort") study, rather than a randomised controlled trial. Snow did not follow each member of the study cohort individually and, in fact, did

Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada
J A Kopec
Department of Medicine, Montreal General Hospital, McGill University, 1650 Cedar Avenue, Montreal, Canada H3G 1A4
J M Esdaile

Correspondence to:
Dr Esdaile

Accepted for publication
April 1990

[a] The term "study base" actually refers to population-time.[11] Some authors also distinguish between a source population, which includes prevalent cases, and a population at risk. The term "study subjects" is used here to refer to the actual cases and controls.

not know the source of drinking water for each household. In order to ascertain the exposure retrospectively, he visited those families in which fatal cases of cholera had occurred. Although Snow obtained the overall number of households supplied by each water company directly from their records, the proportion exposed in the study population could have been estimated by measuring the quality of water in a random sample of all households. This example illustrates that under optimal conditions, neither "backward" directionality nor sampling of the base make the case-control method inherently less valid, or scientifically inferior to other types of research in epidemiology.

The idea of using the experimental approach as a paradigm for observational research was strongly advocated by Hill[15] and, more recently, by Horwitz and Feinstein.[4 16–19] Exactly how this paradigm should be implemented remains problematic. According to Feinstein, special care in the conduct of observational research should be devoted to the selection of subjects according to the same eligibility criteria that are commonly applied in randomised controlled trials (eg, indications, contraindications), and the avoidance of susceptibility bias (confounding by susceptibility), detection bias in ascertaining the cases, protopathic bias (confusing cause and effect) and "transfer bias".[4] Esdaile and Horwitz[20] used the randomised controlled trial framework to develop a list of principles that could be used as a check list for a scientific quality of an observational study. Miettinen[21] pointed out that the traditional outlook in case-control studies is incompatible with the randomised controlled trial paradigm, where the concern is always to compare exposed and unexposed, rather than cases and "non-cases". He also emphasised the need to define appropriately not only exposure but also "non-exposure" (the reference level of exposure), and the importance of restricting the study population, in order to obtain a desired distribution of exposure levels.

The usefulness of the randomised controlled trial paradigm in elucidating potential sources of bias stems primarily from its conceptual simplicity as a model for cause-effect research, and the relatively advanced knowledge of the methodological principles involved. In the following sections of the paper we will apply this paradigm to examine common sources of bias in case-control research. Three broad issues will be discussed: (1) problems related to the definition of the study population; (2) measurement related biases; and (3) bias due to non-participation.

Study population

POPULATION BASED STUDIES

Experimental populations are "fixed" cohorts, with membership defined automatically by the fact of "enrolment" into the study. In "case-control within cohort" studies the population which gives rise to the cases is an observational cohort, and is defined in a similar way. Most of case-control studies are conducted within so called dynamic (open) populations, characterised by a constant influx and outflow of subjects. Nevertheless, a dynamic population can be precisely defined, eg, as the population of a geographical area, employees in an industrial plant etc, over a specified time period.

Once the study population (study "base") has been defined, the validity of the study depends on the investigator's ability to obtain accurate information on the outcome for all its members (as in a randomised controlled trial), and information on the exposure and other relevant variables (confounders, effect modifiers) for all of the cases, and a random sample of the base.[11] It may be noted that in a randomised controlled trial, exposure information is available for all individuals. The selection of controls (referents) in population based studies should follow the principles of random sampling. Various methods of sampling have been proposed: (1) traditional "cumulative" sampling, at the end of the observation period; (2) "case-base" or "case-cohort" sampling, performed at the start of the observation period; (3) "incidence-density" sampling, conducted throughout the period of follow up.[22 23] In order to adjust for confounding factors, the sampling is usually performed within strata, formed by the categories of potential confounders. A useful strategy, especially when the study population is "unstable", is to consider time as one of the stratification variables.[24 25]

HOSPITAL BASED STUDIES

In so called hospital based, or "case initiated" case-control studies,[26] the investigator starts with a case series, usually selected from a medical care facility, and compares it (in terms of exposure history) to an arbitrary control group, taken either from the hospital (patients with other diseases) or from the community (eg, neighbours or relatives of the cases). The cases and controls are usually matched for a number of potentially confounding variables. The source population in these studies is often described vaguely as those subjects who are "at risk" of both exposure and disease,[7] but no effort is made to delineate it with any rigour. As a result, two types of problems arise. First, the investigator has to accept that the cases studied are only a fraction of all cases that might be "eligible" for the study. A substantial and potentially differential underdetection cannot be ruled out, but its extent can hardly be evaluated. Second, there is no "roster" from which to select a population based control series. The resultant problem of potential "selection" bias[a] can be seen as one of "incoherence" of the cases and controls with respect to the population at risk they represent.[6 12]

SECONDARY BASE

Conceptually, some improvement in the design of case initiated studies is possible by "turning the argument around" and defining the study population secondarily, as all those subjects who would have been included in the case series, had they developed the disease of interest.[11] Full case ascertainment is then achieved simply by definition, and the previously mentioned selection bias that arises in the case series is apparently eliminated. However, since the

[a] We try to avoid the term "selection bias" because of its ambiguity and because it has been used in many different contexts.

membership in the population at risk is now defined only conditionally on a hypothetical event, it cannot be determined with certainty. Thus the problem of unbiased sampling of controls remains unresolved.

The choice of an appropriate control group in case initiated (hospital based) studies has been the subject of much controversy. One proposed solution is to take cases of other, carefully selected diseases, from the same "source" as the original case series. To avoid bias, the control disease must not be related to the exposure at issue, and should be subject to exactly the same surveillance and detection procedures as the disease of interest. For many diseases, and some common exposures, these conditions seem to be very difficult to satisfy, and may actually conflict with each other.[7 27 28] Such factors as care seeking behaviour, diagnostic accuracy, referral patterns or hospital catchment areas are likely to differ for different diseases.[4] Using a "diagnostic roster", (eg, a chest *x* ray file) as a source of both cases and controls does not seem to be an improvement. Although the "medical itinerary" of the cases (eg, cases of lung cancer) and the controls (other lung diseases, etc) may be similar, they may often be related to the same exposure studied (eg, smoking). The same objection applies to an extreme version of the above strategy, namely to compare different histological subtypes of the same disease.[29] Another common option in selecting the control group is to use "community controls". This approach can seldom be theoretically justified in case initiated studies.[6 12] The use of neighbours, siblings or friends of the cases has more to do with controlling confounding (through matching), than it does with unbiased sampling of the base. The amount of bias will certainly be study specific, and it is difficult to estimate. Only limited attempts to study the problem empirically have been made.[30] In one study in which friends of the cases have been used as controls,[31 32] the "friendly control bias" apparently increased the observed effect. In other studies it may act in the opposite direction.

As an efficient alternative to the case initiated study some authors employed a "control initiated" design,[33 34] whereby the investigator uses the same control series more than once, to study different diseases and different exposures. This design has been used infrequently, and it does not offer any advantage over the case initiated design in terms of validity.[26]

Because of these difficulties, many epidemiologists (including the authors of this review) believe that validity of the control group is most likely to be achieved, and easiest to evaluate, when the case series can be linked to an identifiable source population. Unfortunately, population based studies are expensive because of the difficulty in ascertaining all the cases. The validity of a hospital (registry, etc) control group depends on the similarity between the control disease and the disease under investigation in terms of all factors bearing on the appearance of the cases in a particular "source", and the lack of association between the control disease and the exposure at issue. Empirical data concerning these properties of different case-control-exposure combinations are scant.

ELIGIBILITY CRITERIA

An internally valid clinical trial can be carried out in a very restricted population, eg, among volunteers, or patients in a single hospital ward. Moreover, the choice of appropriate "eligibility" criteria may significantly improve the study's validity, for example by increasing compliance, minimising drop out rate, or enhancing the quality of measurement.[35] The same principles apply in case-control studies. Failure to apply certain restrictions on admissibility may seriously limit the investigator's ability to control bias.[16–21] The specific admission criteria will usually differ considerably in studies of therapy (most of which are experimental) and studies of aetiology (most of which are observational). For example, in case-control studies on side effects of drugs and other therapeutic manoeuvres, it does not matter whether the subject had or did not have indications for the therapy used, except when a given indication can be considered a confounding variable or an effect modifier.[21] Similarly, it is irrelevant whether the subjects had the "opportunity" for exposure or not.[36] Measurement accuracy is usually a greater problem in case-control studies than in randomised controlled trials and, as will be discussed, restricting the source population is an important strategy in dealing with this problem. In both types of studies it is reasonable to exclude those subjects whose risk of developing the disease of interest is known to be zero.[11 25]

Measurement

MEASUREMENT OF OUTCOME (CASE ASCERTAINMENT)

By analogy with experimental research, it seems useful to approach the problem of case identification in case-control studies from a measurement point of view. In a large clinical trial, one would have an elaborate system of *active surveillance* for all subjects in the study population, perhaps involving periodic follow up examinations. The out migrants would be traced, the subjects' personal physicians might be contacted, and the records reviewed. The diagnosis in each case would be verified by independent, "blinded" reviewers, according to standard, pre-established criteria, and using all pertinent information available.[35]

Unfortunately, maintaining such a system for a long period of time, even in a moderately large population, may be expensive and difficult logistically, as the experience of some clinical trials and observational cohort studies suggest. The alternative approach, commonly employed in case-control studies of rare diseases, is to replace these costly activities with *routine recording* of contacts between patients and the health system (including deaths) as the primary source of data on the outcome. The measurement (case identification) thus becomes a multistage process, largely beyond the investigator's control. For a case to be properly identified, the following activities must take place: (1) a person must receive medical attention; (2) he/she must be properly diagnosed as having the disease of interest; and (3) the record of the diagnosis must appear in the "source" of the case series (eg, a

registry, a hospital, etc). The remaining members of the base are, by default, regarded as "non-cases". Since each missed case represents an instance of measurement error (misclassification), it becomes clear that the objective should always be to ascertain all incident cases that occur in the study population. For the purpose of analysis, a random sample of the cases may sometimes be sufficient. In some studies there will also be persons without the disease, erroneously diagnosed as cases ("false positives"). Misclassification of subjects in terms of their disease status, even if "non-differential" with respect to the exposure, will usually produce a bias toward the null in the estimate of causal effect. (An important exception from this rule will be discussed later.) Differential misclassification may produce bias in either direction.

The objective of complete ascertainment of cases in the study base may be unrealistic. For many conditions the accuracy of medical diagnosis is imperfect. Increasingly, sophisticated and specialised care is necessary to make the diagnosis. Even for conditions that almost invariably lead to serious symptoms and hospital admission, such as many cancers, the completeness of reporting seldom exceeds 95%, and may be much lower.[37] Many relatively mild disorders are never recorded simply because medical attention is not sought. The incident cases of some chronic age related disorders, such as osteoarthritis, are very difficult to ascertain from medical records, even though these conditions produce significant morbidity and disability. In addition, hospital catchment areas and referral patterns are often difficult to specify with precision. Thus, for most diseases, the task of identifying all cases in any large population may appear insurmountable.

There are several strategies in the study design that the investigator may employ to reduce misclassification of the outcome status and avert bias. The diagnosis can often be verified in the cases, through the use of various auxiliary data sources, thus eliminating "false positives". In this way, misclassification is reduced to under-ascertainment of the cases, a situation that may be easier to handle, as discussed below. Examining the controls to detect "false negatives" is recommended by some authors,[4] but may be difficult, especially when sophisticated technology, invasive tests or unpleasant procedures are required to rule out a given diagnosis.

When information is gathered prospectively, the investigator may consider various forms of active case ascertainment, similar to those used in large experimental or observational cohorts. In some cases, a periodic survey of all members of the base may be an option, especially in studies of relatively mild but self evident illnesses (eg, minor accidents). In retrospective studies, multiple data sources should be used to improve detection. For example, to identify all cancers in a given population over a specified time period, the investigator may review the death certificates, hospital records, and pathology reports, in addition to the data from a tumour registry.[37]

Although diagnosis verification, active case ascertainment, and the use of all available data sources may significantly improve validity, substantial and potentially differential underdetection may still remain. It is important to recognise that non-differential under-ascertainment of the cases will not bias the estimate of the rate ratio.[38] Thus another strategy is to ensure, to the extent possible, that all activities associated with case ascertainment are performed with the same vigour among the exposed and unexposed, and follow the same, standard protocol. In randomised controlled trials this is achieved through blinding of the manoeuvre (eg, a placebo) and blind assessment of the outcome. In case-control studies, exposure status is often unknown to the physician making the diagnosis, and "unrealised" by the patient, which would make differential bias unlikely. This situation may be changing, however, as information about newly discovered, or only suspected health hazards is increasingly considered a "hot topic" by the media. When reviewing medical records or verifying the diagnoses, the evaluators should be blind as to whether the subject is exposed or not.

Another useful strategy is to restrict the study population to those who, because of their demographic profile (age, sex, ethnicity, education, occupation, income) or some other characteristic specific to the problem studied (eg, being covered by a particular health plan, living close to the study hospital), have a reasonable chance of appearing in the case series should they develop the disease under investigation. Then, of course, those cases that do not meet the prespecified criteria must be excluded. As part of the criteria for admission, it will often be necessary to modify the definition of a case. For example, only serious cases with overt symptoms are accepted. The study population may have to be redefined once all potential sources of cases have been determined and it turns out that certain segments of the population are not covered. Savitz and Pierce[39] illustrate how these problems might be approached in the case of male infertility, a condition with a very high degree of underascertainment. The study population should be restricted to those men who (1) want to have children and (2) are willing to seek medical care when this desire is unfulfilled. Then, given that appropriate care is devoted to the identification of cases, a high degree of ascertainment may be achieved. At the same time, the selection of a representative control group will be possible, provided that such attributes as "child seeking behaviour" can be validly measured (eg, by means of a questionnaire). In practice, one could initially select a larger pool of potential controls from a "candidate" population and then eliminate those who do not qualify as members of the base ("child seekers"). With the above approach, the distinction between "primary base" and "secondary base" becomes somewhat artificial.[39] The whole process of designing a case-control study is in a sense iterative. The challenge is to define the study population in such a way that all its members (over a specified time period) can be correctly classified as cases or "non-cases", using available case detection procedures.

MEASUREMENT OF EXPOSURE

In a laboratory experiment or, to a somewhat lesser degree, in a randomised controlled trial, the investigator has almost full control over the specific type of agent being investigated, as well as the dose, timing and duration of exposure. None of these elements is controlled by the investigator in a case-control study, in which the estimate of exposure is usually based on past records, made for other purposes, or recall of past events by the subjects. Even the very definition of exposure may be a problem, let alone the measurement of the dose.[40] Of course, the specific techniques used to obtain exposure data will depend on the type of exposure studied. These issues are too complex to be discussed here, but excellent references are available (eg[41]).

In some case-control studies exposure is recorded before the onset of disease, or after the onset of disease but before the assessment of disease status. However, it is ascertained by the investigator after the disease status has been assessed. Thus the possibility of differential misclassification of exposure, often referred to as "recall bias" when the assessment is based on self report, should always be taken into account. Another form of measurement related bias in case-control studies is protopathic (reverse causality) bias. Here, the ascertainment of exposure is affected by disease related changes in the level of exposure.[42]

Perhaps the most common method of data collection in case-control studies is self report. The accuracy of recall in an interview has been studied extensively for a number of common exposures, for example smoking, medication associated with female reproduction, past history of chronic illness and hospital admission, diet, and many occupational exposures. A common finding is that the quality of information depends on both the type of data sought and the methodological aspects of the assessment process.[43–46] For example, memory aids, such as pictorial displays, food models, or wall calendars have been found useful in a number of studies.[43 45 47] In a series of experiments on survey techniques, substantial improvement in reporting, particularly for "sensitive" topics or "socially undesirable" behaviours, was observed when the following techniques were used: (1) a mixture of long and short questions; (2) detailed respondent instructions; (3) feedback (following each question), informing the respondent how well he is performing; and (4) commitment technique, consisting in asking the respondent to sign an agreement to respond conscientiously.[48] The so called cognitive approach to questionnaire design has been extensively studied at the National Centre for Health Statistics, and the preliminary results are encouraging.[49]

In many situations, despite the use of state of the art measurement techniques, significant bias or imprecision in the estimate of exposure is difficult to avoid. If this occurs, an important consideration is whether these measurement errors are likely to be different (in terms of frequency, magnitude or direction) among the cases and the controls. Differential errors are usually considered a greater threat to the study's validity than non-differential ones, since in the former the direction of bias in the estimate of effect may be away from the null and may be difficult to predict.[50] Differential report (recall) bias has been implicated mainly in studies on pregnancy outcomes.[51] There are also examples of recall bias in psychiatric publications.[52] The suggestion that the apparent relationship of soft tissue sarcoma and malignant lymphoma to phenoxy acids may be due to recall bias has not been confirmed.[53] Overall, the empirical evidence concerning the extent and frequency of recall bias is scant.[54] Of two recently reported studies, one failed to provide any evidence of biased reporting of exposure during pregnancy,[55] whereas the second study (using a somewhat different design) suggested the existence of substantial recall bias for some, but not all, exposure factors.[56]

The most effective device the investigator can employ to ensure comparability of measurement is blind assessment of exposure. That is, neither the subject nor the investigator (or interviewer) should know who is a case and who is a control. This is analogous to double blinding, routinely employed in randomised controlled trials. Despite its strong appeal this obvious precaution is sometimes ignored in case-control studies. If blinding is not feasible, it is essential that the same standard measurement conditions and procedures are used for all subjects. In some instances this approach may lead to a conflict between the principle of maximum accuracy and that of maximum comparability of measurement. For example, if data for the cases have been obtained through proxies, it may be advisable to obtain data for the controls in the same way, even if it may cause greater bias (but hopefully similar to that in the cases) or less precision than questioning the controls directly.[12] On the other hand, there is no evidence to support the need for using dead controls for dead cases.[57]

Although the use of some equivalent to a placebo to simulate the disease (or death) among the controls is not feasible in most case-control studies, the investigator may select the controls from subjects with other disorder(s). This method is commonly used in hospital based studies. Disease controls may also be used as a second control group in population based studies, if comparability of information appears to be a significant problem.

It should be emphasised that the methods of controlling measurement bias in the analysis are very limited. Predictors of measurement errors cannot, in general, be treated in the same way as confounders in the analysis, and improved measurement quality is always preferable to analytical adjustments.[58] Nevertheless, it may sometimes be possible to correct algebraically for measurement bias after the data have been collected, if a validation substudy has been performed. Even then, for a wide range of conditions, it may be more efficient to use the optimal (but more expensive) method of measurement in all subjects, and enrol fewer subjects into the study.[59] It has been also suggested that a validity subscale be incorporated into the questionnaire, using "fake" exposures, whose relation to the disease studied has been ruled out.[51] Alternatively, an additional, "objective" source of information, such as

medical records, may be available for some exposures, but not for others; if the two exposures are comparable in terms of validity of recall, the magnitude and direction of bias can be estimated, and an appropriate correction factor can be incorporated into the analysis. This approach has a serious limitation as there is some empirical evidence that different exposures tend to behave differently with respect to reporting bias.[44 45 56]

Sometimes, the best tactic to obtain good quality data is to restrict eligibility to those for whom valid information can be collected. This is somewhat analogous to checking the compliance of potential candidates for a randomised controlled trial. Such restrictions should be made a priori (or, at least, without reference to the outcome data collected in the study), and may involve age, ability to communicate in a given language, mental abilities, educational level, occupational group, etc. For example, when occupation is used as a surrogate measure of exposure, the researcher conducting a case-control study may select the occupational groups with definitely high or low (plus, perhaps, intermediate) exposure levels, and exclude all cases and controls whose occupation is not a good indicator of exposure. Such an approach resembles in some respects a cohort study, and is consistent with the randomised controlled trial paradigm.

Participation

In randomised clinical trials, any confounding potentially introduced by arbitrary exclusions/inclusions, refusal to participate, or other forces operating prior to the moment of randomisation is effectively prevented. However, exclusions, withdrawals or losses to follow up occurring after randomisation are a potential source of "selection" bias.[35 60] Differential losses can often be prevented by true blinding with respect to the exposure. If blinding is not possible, the success with which the follow up is achieved becomes the basic measure of the quality of the study.[8] Yet, relatively little attention has been paid to high participation rate in evaluating the quality of case-control studies.[6] Even if we assumed that no confounding by unknown factors was initially present in the source population (as if randomisation had been performed), the possibility of "selection" bias due to missing data, non-response, incompleteness of medical records, obvious errors in measurement, and other uncontrollable factors should always be taken into account. For example, cancer patients who are aware of having been exposed to a widely publicised risk factor may be more likely to participate than healthy or unexposed individuals. A positive effect may then be found even if in fact there is no association between the disease and the exposure. This type of bias may occur if the prospective participants (members of the source population) are aware of the exposure and disease being investigated, and the response rate is low. Exposure differential non-participation among the controls will bias the estimate of the proportion exposed in the source population.

Participation bias should not occur if the relevant information is obtained for all subjects initially selected for the study, and all exclusions are based on a priori defined criteria (definition of the base). In practice, one should probably strive for a participation level similar to that required in other types of studies, ie, 80–90% in both cases and controls. This will sometimes call for the use of special techniques, developed by social scientists and large survey organisations. It has been found, for example, that availability of potential respondents in a population survey depends on the timing and number of calls made by the interviewer, whereas refusals to participate are related to, among other things, fear of crime and concerns about confidentiality.[61] When data are collected by means of a self administered questionnaire, the method known as "total design method" may considerably improve the response rate.[62] In some studies, the low response rate was due to such simple omissions in the planning of a study as the investigator's failure to print a sufficient number of questionnaires for a second and third mailing.[62] Whenever feasible, the subjects should be kept blind as to the hypothesis under investigation. Restricting the study population to those subjects who are likely to participate in any study, regardless of the hypothesis, may be a very useful strategy and should be considered whenever such general predictors of cooperation are known beforehand and are easy to measure.

Conclusions and recommendations

It seems reasonable to argue that under optimal conditions, a case-control study is as valid as a randomised clinical trial. For many problems, a nested study provides evidence as strong as a corresponding cohort study, but may be much more efficient. This is particularly true when the information on exposure is obtained from past records. Of course, confounding remains a critical issue in observational research and neither careful subject selection nor measuring known confounders, followed by adjustment in the analysis, will ever provide the level of comparability of the exposed and unexposed groups that is attainable through randomisation. This weakness of the observational design in studying certain behavioural interventions has been elegantly demonstrated by Grey-Donald and Kramer in their study of the relationship between formula supplementation in hospital and duration of breast feeding.[63]

In this review we have deliberately omitted the topic of confounding, and focused on other sources of bias, particularly relevant in the context of case-control studies. The ideas presented here are not new, and have been discussed in a number of recent texts and articles on methods in epidemiology. Rather than emphasise differences between the authors, we have attempted to reconcile and bring together various seemingly divergent points of view. Our point of departure, and a unifying theme, was the conceptual link between a case-control study and a randomised clinical trial. A number of bias minimising strategies have been discussed and these are summarised below.

(1) In all case-control studies, an attempt should be made to define explicitly the source population for the cases.

(2) If the case series can be linked to an identifiable source population, a valid control group can be obtained by random sampling.

(3) If hospital (registry, etc) controls are used, the control disease(s) should be subject to the same case detection mechanisms/procedures as the disease of interest, and should not be related to the exposure at issue.

(4) In population based studies, complete ascertainment of all cases should be sought. This can be facilitated by: (a) a suitable definition of the outcome; (b) an extensive use of multiple data sources and active case detection; and (c) restricting the source population to those presumably covered by the available ascertainment systems/procedures.

(5) Additional eligibility criteria can be applied, in order to improve the quality of information and prevent participation bias.

(6) State of the art methods of measurement should be used, including sophisticated questionnaire designs.

(7) Blind assessment is the best way to ensure comparability of information in the cases and in the controls. If blinding is not feasible, standardised, uniform data collection procedures must be used in all subjects.

(8) The subjects and interviewers should not be informed about the hypothesis under study. Diagnosis verification should be done "blindly".

(9) High response rate among both cases and controls should be obtained.

The authors are grateful to Drs J-F Boivin, R I Horwitz and A B Miller for their constructive criticism and helpful comments on preliminary drafts of this manuscript. However, the views expressed in the article are those of the authors and do not necessarily reflect the views of Drs Boivin, Horwitz, or Miller.

Dr Kopec is the recipient of a Fellowship from the National Health Research and Development Program of Canada. Dr Esdaile is a Senior Research Scholar of the Fonds de la recherche en santé du Québec. Supported in part by grants from the Arthritis Society of Canada and the Institut de recherche en santé et sécurité du travail.

1 Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: principles and quantitative methods*. Belmont: Wadsworth, 1982.
2 Rothman KJ. *Modern epidemiology*. New York: Little, Brown, 1986.
3 Sackett DL. Bias in analytic research. *J Chron Dis* 1979; **32**: 51–63.
4 Feinstein AR. *Clinical epidemiology. The architecture of clinical research*. Philadelphia: WB Saunders, 1985.
5 Last JM, ed. *A dictionary of epidemiology*, 2nd ed. Oxford: Oxford University Press, 1988.
6 Cole P. Introduction. In: Breslow NE, Day NE. *Statistical methods in cancer research. Vol 1. The design and analysis of case-control studies*. Lyon: IARC, 1980.
7 Schlesselman JJ. *Case-control studies: design, conduct, analysis*. New York: Oxford University Press, 1982.
8 Breslow NE, Day NE. *Statistical methods in cancer research. Vol 2. The design and analysis of cohort studies*. Lyon: IARC, 1987.
9 Kramer MS, Boivin J-F. Toward an unconfounded classification of epidemiologic research design. *J Chron Dis* 1987; **40**: 683–8.
10 Greenland S, Morgenstern H. Classification schemes for epidemiologic research design. *J Clin Epidemiol* 1988; **41**: 715–6.
11 Miettinen OS. *Theoretical epidemiology. Principles of occurrence research in medicine*. New York: John Wiley & Sons, 1985.
12 Miettinen OS. The case-control study: valid selection of subjects. *J Chron Dis* 1985; **38**: 543–8.
13 Mantel N. Synthetic retrospective studies and related topics. *Biometrics* 1973; **29**: 479–86.
14 Snow J. *On the mode of communication of cholera*. 2nd ed. London: John Churchill, 1960.
15 Hill AB. Observation and experiment. *N Engl J Med* 1953; **248**: 995.
16 Horwitz RI, Feinstein AR. The application of therapeutic-trial principles to improve the design of epidemiologic research: a case-control study suggesting that anticoagulants reduce mortality in patients with myocardial infarction. *J Chron Dis* 1981; **34**: 575–83.
17 Feinstein AR. Experimental requirements and scientific principles in case-control studies. *J Chron Dis* 1985; **38**: 127–33.
18 Feinstein AR. Epidemiologic analyses of causation: the unlearned scientific lessons of randomized trials. *J Clin Epidemiol* 1989; **42**: 481–9.
19 Horwitz RI. The experimental paradigm and observational studies of cause-effect relationship in clinical medicine. *J Chron Dis* 1987; **40**: 91–9.
20 Esdaile JM, Horwitz RI. Observational studies of cause-effect relationships: an analysis of methodologic problems as illustrated by the conflicting data for the role of oral contraceptives in the etiology of rheumatoid arthritis. *J Chron Dis* 1986; **39**: 841–52.
21 Miettinen OS. The clinical trial as a paradigm for epidemiologic research. *J Clin Epidemiol* 1989; **42**: 491–6.
22 Kupper LL. McMichael AJ, Spirtas R. A hybrid epidemiologic study design useful in estimating relative risk. *J Am Stat Assoc* 1975; **70**: 524–8.
23 Miettinen OS. Design options in epidemiologic research. An update. *Scand J Work Environ Health* 1982; **8** (Suppl): 7–14.
24 Breslow N. Design and analysis of case-control studies. *Annu Rev Public Health* 1982; **3**: 29–54.
25 Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *Am J Epidemiol* 1982; **116**: 547–53.
26 Greenland S. Control-initiated case-control studies. *Int J Epidemiol* 1985; **14**: 130–4.
27 Axelsson O. Comment. The "case-control" study: valid selection of subjects. *J Chron Dis* 1985; **38**: 553–5.
28 Axelsson O, Flodin U, Hardell L. A comment on the reference series with regard to multiple exposure evaluations in a case-reference study. *Scand J Work Environ Health* 1982; **8** (Suppl): 15–19.
29 Knottnerus JA. Letter to the Editors. Subject selection in hospital-based case-control studies. *J Chron Dis* 1987; **40**: 183–5.
30 Stavraky KM, Clarke EA. Hospital or population controls? An unanswered question. *J Chron Dis* 1983; **36**: 301–7.
31 Siemiatycki J, Colle S, Campbell S, Devar R, Belmonte MM. Case control study of insulin dependent (type I) diabetes mellitus. *Diabetes Care* 1989; **12**: 209–16.
32 Siemiatycki J. Friendly control bias. *J Clin Epidemiol* 1989; **42**: 687–8.
33 Weiss NS, Daling JR, Chow WH. Incidence of cancer of the large bowel in women in relation to reproductive and hormonal factors. *J Natl Cancer Inst* 1981; **67**: 57–60.
34 Harris NV, Weiss NS, Francis AM, Polissar L. Breast cancer in relation to patterns of oral contraceptive use. *Am J Epidemiol* 1982; **116**: 643–51.
35 Peto R, Pike MC, Armitage P, *et al*. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976; **34**: 585–612.
36 Poole C. Exposure opportunity in case-control studies. *Am J Epidemiol* 1986; **123**: 352–8.
37 Robles SC, Marrett LD, Clarke EA, Rish HA. An application of capture–recapture methods to the estimation of completeness of cancer registration. *J. Clin Epidemiol* 1988; **41**: 495–501.
38 Poole C. Exceptions to the rule about nondifferential misclassification (Abstract). *Am J Epidemiol* 1985; **122**: 508.
39 Savitz DA, Pearce N. Control selection with incomplete case ascertainment. *Am J Epidemiol* 1988; **127**: 1109–17.
40 Matanoski GM. Issues in the measurement of exposure. In: L Gordis, ed. *Epidemiology and health risk assessment*. Oxford: Oxford University Press, 1988.
41 Gordis L, ed. *Epidemiology and health risk assessment*. Oxford: Oxford University Press, 1988.
42 Horwitz RI, Feinstein AR. Methodologic standards and contradictory results in case-control studies. *Am J Med* 1979; **66**: 556–64.
43 Mitchell AA, Cottler AB, Shapiro S. Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 1986; **123**: 670–6.
44 Harlow SD. Agreement between questionnaire data and medical records. The evidence for accuracy of recall. *Am J Epidemiol* 1989; **129**: 233–48.
45 Lee-Han H, McGuire V, Boyd NF. A review of the methods used by studies of dietary measurement. *J Clin Epidemiol* 1989; **42**: 269–78.

46 Jain M, Howe GR, Harrison L, Miller AB. A study of repeatability of dietary data over a seven year period. *Am J Epidemiol* 1989; **129**: 422–9.
47 Beresford SAA, Coker AL. Pictorially assisted recall of past hormone use in case-control studies. *Am J Epidemiol* 1989; **130**: 202–5.
48 Cannel CF, Miller PV, Oksenberg L. Research on interviewing techniques. In: Lainhard S, ed. *Sociological methodology*. San Francisco: Jossey-Blass, 1981.
49 Jobe JB, Mingay DJ. Cognitive research improves questionnaires. *Am J Public Health* 1989; **79**: 1053–5.
50 Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977; **105**: 488–95.
51 Raphael K. Recall bias: a proposal for assessment and control. *Int J Epidemiol* 1987; **16**: 167–70.
52 Wolkind S. Koleman EZ. Adult psychiatric disorder and childhood experiences: the validity of retrospective data. *Br J Psychiatry* 1983; **143**: 188–91.
53 Hardell L. Relation of soft tissue sarcoma, malignant lymphoma and colon cancer to phenoxy acids, chlorophenols and other agents. *Scand J Work Environ Health* 1981; **7**: 119–30.
54 Lippman A, MacKenzie SG. What is recall bias and does it exist? In: Marois M, ed. *Progress in clinical and biological research*, vol 163. New York: Alan R Liss, 1985.
55 MacKenzie SG, Lippman A. An investigation of report bias in a case-control study of pregnancy outcome. *Am J Epidemiol* 1989; **129**: 65–75.
56 Werler MM, Prober BR, Nelson K, Holmes LB. Reporting accuracy among mothers of malformed and nonmalformed infants. *Am J Epidemiol* 1989; **129**: 415–21.
57 Gordis L. Should dead cases be matched to dead controls? *Am J Epidemiol* 1982; **115**: 1–5.
58 Greenland S. Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985; **122**: 496–506.
59 Greenland S. Statistical uncertainty due to misclassification: implications for validity substudies. *J Clin Epidemiol* 1988; **41**: 1167–74.
60 Armitage P. Exclusions, losses to follow-up, and withdrawals in clinical trials. In: Shapiro SH, Louis TA, eds. *Clinical trials. Issues and approaches*. New York: Marcel Dekker, 1983.
61 Hawkins DF. Nonresponse in Detroit area study surveys: a ten year analysis. Working papers in methodology No 8. Chapel Hill, N C: Institute for Research in Social Science, 1977.
62 Dillman DA. Mail and other self-administered questionnaires. In: Rossi PH, Wright JD, Anderson AB. *Handbook of survey research*. New York: Academic Press, 1983.
63 Grey-Donald K, Kramer MS. Causality inference in observational vs. experimental studies. An empirical comparison. *Am J Epidemiol* 1988; **127**: 885–92.