British Journal of Psychiatry (1990), **157**, 197–207

The Case-Control Study in Psychiatry

GLYN LEWIS and ANTHONY J. PELOSI

Case-control studies involve 'cases' being compared to 'controls' with respect to 'exposures', possible aetiological (or associated) factors. Associations between a disease and an exposure can be explained by chance, reverse causality, confounding and biases or, lastly, by causality. However, confounders as well as information and selection biases can be adjusted for at the design (or analysis) stage of the study. The strength of an association can be measured by means of relative risk, calculated indirectly using the odds ratio. Well conducted control studies should produce accurate estimates of relative risks in many psychiatric investigations.

"In a case-control study, individuals with a particular condition or disease (the *cases*) are selected for comparison with a series of individuals in whom the condition or disease is absent (the *controls*). Cases and controls are compared with respect to existing or past attributes or exposures thought to be relevant to the development of the condition or disease under study" (Schlesselman, 1982).

Case-control studies thus estimate the strength of association between a disease and a factor of interest and are widely used as a quick, efficient, and relatively inexpensive means of exploring possible aetiological associations. Mantel & Haenszel (1959) observed that the method is "adapted to the limited resources of an individual investigator and places a premium on the formulation of hypotheses for testing, rather than on facilities for data collection".

The refinement of this methodology has been one of the major advances in epidemiological research over the past three decades. Many of the technical developments have been by epidemiologists interested in the aetiology of cancer who have done much to remove the disreputable image of the 'quick and dirty' case-control study that provided unreliable results requiring confirmation with other study designs. The case-control study now has a central respected place in aetiological research.

Its application to mental disorders will also involve those particular problems which bedevil the psychiatric researcher. However, psychiatry is only a special case within medicine and the similarities between our problems and those of our medical colleagues are more striking than the differences. Where possible we shall attempt to discuss the particular problems of psychiatry and how they have been or could be overcome.

This paper is intended to help those planning studies as well as those who want to understand more about interpreting the results of scientific papers. Also, professional bodies in all areas of medicine, including psychiatry, are stressing the need for a more thorough knowledge of epidemiology and experience of research.

There is insufficient space to explain fully all the topics covered here, but reference will be made to more detailed accounts. General problems of study design, interpretation and confounding will be outlined, before discussing those difficulties which raise particular problems in case-control studies: selection bias and information bias.

Nomenclature

Exposure

Epidemiologists tend to use 'cases' to refer to cases of the disease or condition and 'exposure' to describe the possible aetiological (or associated) factor of interest, although in many circumstances the latter term is clearly inappropriate. 'Exposures' will be broadly defined and will not be restricted to a history of past contact with, for instance, a noxious chemical, infection or social adversity. A current physiological characteristic or an ongoing or long past contact with a possible causal factor will all be called 'exposures'. Case-control studies have thus been used to study a range of psychiatric disorders and exposures, such as the association between schizophrenia and computerised tomography (CT) scan parameters (e.g. Smith *et al*, 1988) and obstetric complications (e.g. Lewis & Murray, 1987), risk factors for tardive dyskinesia (e.g. Mukherjee *et al*, 1986) and depression and corticosteroid metabolism (e.g. Carroll *et al*, 1981). Geneticists who study the association of a disorder with an increased familial risk are also using a case-control design (e.g. Coryell & Tsuang, 1982).

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press

This illustrates that epidemiology crosses the ideological boundaries of psychiatry, and its methods are as essential for those interested in biological factors as for social psychiatrists. There is nothing more 'biological', for instance, than the link between smoking and lung cancer, and this association was discovered using a case-control study (Doll & Hill, 1952).

Types of study

The nature of case-control studies will be clarified by comparing them with other designs. In the past, the terms 'prospective' and 'retrospective' have been used to describe cohort and case-control studies but these terms are sometimes used to refer to the way in which the data was collected rather than the logical basis of the study. Hopefully the following classification of types of study will avoid this ambiguity.

(1) Clinical description

Describing a series of cases of a disease is still a useful and important research activity. However, the association between a condition and an exposure is rarely so strong that a clinical description provides enough evidence for it. For less clear-cut associations, a comparison group is needed in order to estimate the prevalence of the history of exposure in the general population.

(2) Case-control studies

The case-control study can therefore be seen as an extension of the traditional clinical description of cases. The proportion of cases giving a history of exposure is compared with that in the controls. The control group gives an estimate of exposed proportion in the population from which the cases were drawn. The researcher therefore usually starts by selecting a source of cases and then chooses an appropriate control group. This is often the most difficult step in designing a case-control study (discussed later). Although 'control' is not very appropriate to describe members of the comparison group, alternatives have not gained widespread acceptance.

(3) Cross-sectional surveys

A population is screened for both the disease and the exposure, allowing the association to be studied. In cross-sectional surveys the cases are identified by means of a population-based survey, and cases can be compared with the non-diseased subjects in the study. Therefore, cross-sectional surveys overcome the important problem of selection bias (see below). However, it has the disadvantage that for uncommon conditions, large numbers of subjects have to be screened in order to detect enough cases.

(4) Cohort or longitudinal studies

The researcher selects two groups without the disease but who differ in terms of their exposure to the possible aetiological factor. The subjects are followed up and the incidence of disease in the two groups compared. For rare diseases this will obviously involve following up very many subjects, but with the more common minor psychiatric disorders, this kind of study can be feasible.

(5) Clinical trials and other experimental designs

It is rarely ethical to study aetiology with an experimental design, although treatment response can sometimes provide clues to the cause of disorder. The key feature which differentiates this type of study is that the investigator randomly allocates subjects (or occasionally clinics or factories) to two or more groups which are treated in different ways. This design randomly allocates all confounding variables, including those of which the researcher is unaware. Confounding can still occur, but does so by chance, and the probability of this is statistically estimated. The main problems in interpretation arise from drop-out from treatment, bias in measurement and uncertainty about generalising the findings. The latter problem and the expense of clinical trials have led some authors to advocate using case-control designs to study treatment (Smith *et al*, 1984).

Choosing a study design in uncommon conditions

Case-control studies have a particular advantage in investigating uncommon illnesses. For instance, the incidence of schizophrenia (less than 20 per 100 000 per year) means that over 500 000 recruits would be required to expect 100 new adult cases in a year's follow-up. For this reason case-control studies, which start with identified cases of schizophrenia rather than healthy subjects, are much more feasible than cohort studies (but see Andreassen *et al* (1987) for an example of the latter).

Studying and interpreting associations

Case-control studies, cross-sectional surveys and cohort studies can measure associations between a disease and exposure and by such studies researchers attempt to study disease aetiology. An associated exposure may be referred to as a 'risk factor', which

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press

can sometimes, but not invariably, imply causality. There are well documented difficulties in inferring cause from an observed association and guidelines to assist these inferences (Hill, 1965). There are four main explanations to be considered when an association is observed.

(1) Chance

An association can arise by chance, and statistical tests of significance estimate the probability of this occurring (the '*P*' value). This means that a result significant at the 5% level will, on average, be found once in every twenty studies even if no true association exists. When this occurs it is known as a type one error and is a particular problem if repeated significance tests are performed within a study.

Less commonly recognised is the opposite error, that a real association is missed because a study is too small and the variation correspondingly large; the study therefore lacks statistical power. Such type two errors are a particular problem in biological studies in psychiatry because the investigations are expensive and sample sizes are therefore very small.

Calculating confidence intervals provides more information than a single '*P*' value (Gardner & Altman, 1986) because it indicates the statistical precision of an estimate as well as the probability of the result occurring by chance.

(2) Reverse causality

This refers to the observation that the disease may cause the exposure. This is often a problem in case-control studies because data on exposure are usually collected retrospectively. For instance marital discord may be associated with depression because depression leads to marital problems. Likewise, low social class, which is associated with schizophrenia, is widely held to result from the illness rather than be a cause of it (Goldberg & Morrison, 1963). In case-control studies therefore, the timing of the exposure and onset of disease is of importance. Similarly, in life-events research much attention is paid to 'independent' life events as opposed to those which may be influenced by the illness under study (Brown & Harris, 1978). In comparison, cohort studies usually start by identifying exposed subjects who do not, at that stage, have the disease, and so reverse causality is less of a problem.

Reverse causality can be an explanation of interest. Some authors have suggested that the association between results of the dexamethasone suppression test and depression has aetiological significance. In contrast, Mullen *et al* (1986) have suggested that changes in corticosteroid metabolism result from the anorexia and insomnia of depression rather than being a possible cause.

(3) Confounding and bias

These are the most important of the possible interpretations of an association. Adjusting for confounding in design and analysis is one of the major preoccupations of medical research. A confounder is an independent risk factor (or protective factor) for the disease at each level of the exposure and is also associated with the exposure. It is not therefore a variable on the causal pathway. A confounding factor can lead to a spurious association or can eliminate a real association between exposure and disease (see below).

There are many potential sources of bias in scientific investigations and the most important of these in case-control studies, selection and information bias are discussed later.

(4) Causality

The last explanation for an association is that the exposure does increase the risk of disease, that it is a cause or on the causal pathway.

Inferring a cause from an association

The main hindrance in making the leap from association to aetiology concerns the problem of confounding. Geoffrey Rose's aphorism that "one cannot account for the confounding factor one does not know about" is a useful reminder when designing studies.

Random allocation of individuals to experimental groups reduces the problem of confounding but people are not randomly allocated to possible causes of disease. Observational studies (case-control, cohort studies and cross-sectional surveys) remain the only way to study environmental causes, and so the problem of confounding is introduced. In attempting to interpret an association it is often helpful to ask "'what are the determinants of exposure in this group, why were some exposed others not?'. In other words, 'How does this study situation differ from a randomised trial?'" (Rose, personal communication).

The final conclusion depends upon judgement, argument and the uses to which the conclusion will be put. In studies which bear directly upon important areas of social policy, one would expect more convincing evidence to be provided by the researcher. Laboratory work provides plausible final pathways for causes, thus supplementing epidemiological

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press

studies that indicate the importance of causal factors in the population at risk.

Incident or prevalent cases

In studying aetiological associations it is recommended that, where possible, new incident cases of disease are selected. If prevalent cases are studied there will be an over-representation of subjects with a poor prognosis who have had the condition for a long period. Therefore an observed association may indicate risk factors for a poor prognosis rather than for the onset of disease.

Size of effect v. statistical significance

The usual method of measuring the strength of an association in aetiological studies is by means of relative risk, i.e. the increased incidence rate of disease in the exposed group expressed as a multiple of the rate in the unexposed group. Cohort studies can estimate the relative incidence rates directly, but in case-control studies this can only be done indirectly using the odds ratio, a simple calculation illustrated in the Appendix.

The use of relative risk gives a measure of the strength of an association and not its statistical significance. This distinction is especially important because the strength of an association is a pointer to aetiological importance (Hill, 1965). For instance, cigarette smoking increases the risk of lung cancer about 10 times, and increases the risk of ischaemic heart disease by 2.5 times. In general, most epidemiologists regard a relative risk above 1.5 or 2 as a risk factor of aetiological interest. Estimates of the relative risk of association between recent life events and depression are around five. Winter birth appears to increase the risk of schizophrenia by less than 1.1 (Eaton, 1985).

Precision: statistical power and study size

The statistical power gives the probability that a type two error will *not* occur. This is more likely to occur when the estimate produced by a study is imprecise. The main method of increasing statistical power is to increase the sample size. In small studies which produce statistically insignificant results a type two error should be considered.

The power of a study depends upon four factors: (1) the strength of the expected association; (2) the prevalence of exposure; (3) the significance level, usually taken as 5%; (4) the sample size. For example, a case-control study looking for a relative risk of two at the 5% level where the prevalence of exposure is 15% has a power of 80% when there are 200 cases and 200 controls.

Unless the researcher is expecting a very strong association between the disease and exposure, a case-control study needs to be quite large to have sufficient statistical power. Although the case-control approach usually needs less resources than cohort studies or surveys, by current standards within psychiatry, statistically powerful case-control studies would still be considered large investigations.

An example of a case-control study

Craig & Brown (1984) used the case-control method to study the possible importance of life events in the aetiology of painful gastrointestinal (GI) conditions. They identified those suffering from 'functional' illness in a GI clinic and compared them with a group from the same clinic who had organic GI disorder and with another group of healthy people from the local community. Table I has been constructed from just some of their results; an environmental stress is defined as a severe life event or major difficulty in a 38-week period preceding the interview.

The efficiency and relative cheapness of the design is clear. If Craig & Brown had designed a cross-sectional community survey to test their hypothesis they would have had to interview many thousands of people before a sufficient number with gastrointestinal disorders would be identified.

In asking whether these life events are of causal significance in functional bowel symptoms the other possible explanatory factors (chance, reverse causality and confounding and bias) must first be considered. The difference in the cases and controls may have arisen by chance but this is most unlikely, as indicated by the usual tests of statistical significance. The 95% confidence limits for the odds ratio are from 2.8 to 13.1. Functional bowel illness could conceivably lead to a greater number of life events (i.e. reverse causality), perhaps mediated by their effects on work performance, relationships, etc. However, the assessment plays close attention to

TABLE I

Life events and painful gastrointestinal (GI) disorder (results adapted from Craig & Brown (1983))

	Functional GI disorder	*Organic GI disorder*
Environmental stress	53	26
No environmental stress	14	42
		Odds ratio = 6.1

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press

timing and independence. Sex and social class, for example, are possible confounders and a number of possible biases have to be considered. Selection and referral biases appear to be the most important. For example, are those with functional bowel symptoms more likely to be referred for hospital treatment if they have recently had life stresses and does this apply to the same extent to the hospital controls?

Confounding

In any observational research the true association between disease and hypothesised exposure can be distorted by confounding variables. Confounders can either lead to a spurious association or eliminate the association between a real risk factor and the disease. Confounding therefore results in a biased estimate of relative risk. If it can be measured then the analysis can be adjusted to take the confounder into account. This makes consideration of confounding essential at the design stage of a study.

A simple illustration is provided by the observation that more women have Alzheimer's dementia than men. This results from two other associations. Firstly, there is a relationship between age and Alzheimer's disease (that it becomes commoner with age) and secondly, women are more likely to live longer than men. Thus age is said to confound the relationship between sex and Alzheimer's disease. Age is a common confounder and needs to be taken into account in almost any study.

Why and how confounding can occur is difficult to explain and is best illustrated using a hypothetical numerical example (Table II) which has been deliberately designed with unrealistically large associations. Suppose a study is attempting to measure the association (if any) between life events and schizophrenia. In the raw data, illustrated in Table II*a*, there is no association, with an odds ratio of one. However, schizophrenics are less likely to have children at home, and if we further suppose that life events are commoner in households that have children, the presence of children in the home confounds the relationship between schizophrenia and life events. When the results are stratified by the presence or absence or children (Table II*b*), there now appears to be a strong association previously masked by the confounder with an overall odds ratio of around five.

A confounder is associated with the disease at each level of the exposure. This is illustrated in Table II*c*, as the data are now stratified by the exposure, life events. The odds ratios now illustrate the association between the confounder (the presence of children) and the disease. The odds ratio of 0.05 in this hypothetical example indicates that schizophrenics are about 20 times *less* likely to have children at home than the controls. For a factor to confound a relationship it must also be associated with the exposure, both within the cases and within the controls. Stratifying the results by the case/control distinction (Table II*d*) illustrates a strong association between life events (the exposure) and the presence of children (the confounder) in this imaginary example.

TABLE II*a*
A hypothetical example of confounding: the relationship between schizophrenia and life events

	Cases	*Controls*
Life event	100	100
No life event	100	100
Total	200	200
	Odds ratio = 1.0	

TABLE II*b*
Stratifying the confounder: the presence of children at home

	Children		*No children*	
	Cases	*Controls*	*Cases*	*Controls*
Life event	45	95	55	5
No life event	5	50	95	50
Total	50	145	150	55
	Odds ratio = 4.7		Odds ratio = 5.8	

TABLE II*c*
Stratifying by the exposure: life events

	Life events		*No life events*	
	Cases	*Controls*	*Cases*	*Controls*
Children	45	95	5	50
No children	55	5	95	50
Total	100	100	100	100
	Odds ratio = 0.04		Odds ratio = 0.05	

TABLE II*d*
Stratifying by the disease: cases or controls

	Cases		*Controls*	
	LE+[1]	*LE−*[2]	*LE+*[1]	*LE−*[2]
Children	45	5	95	50
No children	55	95	5	50
Total	100	100	100	100
	Odds ratio = 15.5		Odds ratio = 19	

1. LE+, life events.
2. LE−, no life events.

Adjusting for confounders

Adjusting for confounders can occur in the design and/or the analysis but is only effective if they are

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press

measured accurately. Inaccurate measurement may result in a false sense of security, since residual confounding may occur despite the presence of a design or analysis that has apparently taken account of it. This can also occur if what is actually measured is an imperfect proxy for the true underlying confounder, for instance, if social class is used as a proxy for differences in lifestyle.

Often many (or even all) of the following four main methods of adjusting for confounding will be used in the same study.

(1) In the design – restriction

Subjects are only selected if they have the same value of the confounding variable. For instance, if sex is the confounder then only women may be selected for study (Brown & Harris, 1978). Jenkins (1985) investigated the sex difference in minor psychiatric morbidity by restricting her sample to one grade of the civil service thus reducing the confounding effects of occupational status and education. Controlling for confounding by restriction however makes it impossible to look for interactions (see below).

(2) In the design – matching

Each case can be matched with controls on the basis of any number of potential confounders, commonly age and sex. By matching, the investigator will have deliberately selected controls who, it is known, are different in regard to exposure from that found in a representative sample of the study population. This means that the matching must be retained in the analysis otherwise the estimate of relative risk may be biased towards unity. Matching can therefore *introduce* bias unless a matched analysis is performed.

Matching can either be on an individual basis or controls can be selected so that the proportion of those with the confounder in the controls conforms to that found in the cases. This is known as frequency matching. There are disadvantages and advantages to both methods (Rothman, 1986).

The main advantage of matching is that it increases statistical power. This is often useful if the disease is rare and cases are difficult to find. In addition, multiple controls for each case can be chosen, although there is usually little benefit in selecting more than four controls per case. However, individual matching can be unnecessarily time-consuming and expensive if cases are matched on several variables because many potential controls may have to be screened in order to find suitable ones.

Another disadvantage occurs if attempts to match on a potential confounder leads to a partial matching on the exposure under study. This is called over-matching because it *reduces* the power of the study.

(3) In the analysis – stratification

The principle of a stratified analysis is straightforward. The odds ratio is calculated within each level of the confounding variable and a summary statistic, a weighted average of some sort, is calculated. The details of analysis will not be described here, but the principle is illustrated in the hypothetical example given above (Table II*a* & II*b*). With stratification the relative risk of disease at each level of the confounding variable cannot be affected by the confounder.

(4) In the analysis – modelling techniques

It is often difficult to control for more than one or two confounders by stratification because the number of subjects in each stratum becomes very small or non-existent. For example, if the experimenter attempts to stratify by sex, age (in six age groups), and social class (six groups) there would be $2 \times 6 \times 6 = 72$ strata. Even with a large study there may be few or no cases or controls in certain strata and so those strata will not be of value in estimating relative risk.

Multivariate analysis is conceptually similar to stratification but makes a simplifying assumption which increases the apparent power of the study (Dunn, 1981). Case-control studies are usually analysed using logistic regression which assumes a linear relationship between the logarithm of the odds (logit) of being a case and the exposures. This model can therefore take account of a large number of variables simultaneously, including continuous measures.

However, there are problems with this method. It assumes a linear (or logistic) relationship (Warr, 1987, discusses non-linear relationships in mental health) and distances the investigator (and the reader) from the data being analysed. It is wise then to use linear modelling in conjunction with simpler stratified analysis.

Interactions or effect modification

Some factors may influence the size of an effect in a different way from that produced by a confounder. Interaction or effect modification occurs when the value of an exposure alters the size of the association between the disease and another risk factor. This is illustrated in Table III with results from a community survey (Bebbington *et al*, 1984). The study was examining the relationship between life events and

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press

TABLE III
An example of an interaction from Bebbington et al *(1984)*

	Working class		*Middle class*	
	Cases	*Controls*	*Cases*	*Controls*
Life events	15	18	4	15
No life events	8	40	8	33
	Odds ratio = 4.17		Odds ratio = 1.10	

the onset of depression. Table III shows the results when they have been stratified by social class. It appears that the association between depression and life events is only apparent in the working-class group. Social class is therefore described as an effect modifier, or alternatively that there is an interaction between life events and low social class in increasing the likelihood of developing depression.

It is useful to compare Table III with the hypothetical example given in Table II*b*. In the latter, when the results are stratified by the presence of children, the odds ratios between life events and the disease are quite similar in the strata – there is no interaction.

Selection bias

Epidemiology is the population-based study of disease and the case-control study is no exception to this rule. The rationale of the case-control design is that the control group should give an estimate of the prevalence of the *exposure in the population from which the cases have been drawn*. Simple algebra illustrates that only under these circumstances will the odds ratio derived from a case-control study conform with the 'true' odds ratio in the study population (Schlesselman, 1982, p. 36). For this reason choosing the control group becomes one of the major problems in this study design. Selection bias occurs when the controls give a biased estimate of the prevalence of exposure in the study population.

Recently, it has been suggested that selection bias may be important in explaining the different results found when examining the association between enlarged cerebral ventricles and schizophrenia (Smith *et al*, 1988). These authors suggested that the variation between studies could be explained by differences in the control groups rather than from differences among the samples of schizophrenic patients.

The issue of selection bias makes the response rate of the study an issue. If a large proportion of the original sample of either cases or controls cannot be contacted then it is possible that the characteristics of those included in the final study will not be representative of the study population. Selection bias will only result if the likelihood of contacting subjects depends on the exposure.

Choosing cases

One consideration is that possible cases should be referrable to a study population. For instance, in a specialised hospital, restricting cases to those of the catchment area may make it easier to choose an appropriate control group.

Choosing a control group

A useful guide for choosing controls is that a control should become part of the sampling frame for the cases if he/she were to develop the disease. For instance, if the cases are all first-contact schizophrenics from a defined catchment area, it would be reasonable to choose a population-based sample from the catchment area as controls. This assumes that everyone with schizophrenia in the catchment area eventually contacts the psychiatric services.

The situation with affective disorder is quite different. There is considerable literature illustrating that the majority of those with affective disorder never contact the psychiatric services (Shepherd *et al*, 1966; Goldberg & Huxley, 1980). Therefore, general population controls for hospital cases of depression could lead to an important selection bias (see example below).

The sampling frame for controls will be chosen partly for practical reasons but will mainly depend on the cases under study and the hypothesis being explored. Several possible sources of controls are considered here.

(1) Clinic or hospital controls

The control group often consists of patients attending the same facility who have been diagnosed as suffering from a psychiatric disorder different from that under study. These subjects provide an unbiased estimate of the exposure in the study population, so long as the exposure is not associated with the control illness and the probability that controls are selected is not influenced by exposure.

It is desirable to avoid using control subjects with one condition. Thus, if there is an association between one control illness and the exposure, a large discrepancy in the relative risk estimate will be avoided.

(2) General population controls

General population controls are frequently used. If a serious psychiatric illness is under study and it is known that all, or almost all, cases in a particular area come to the attention of the local clinic, a general population control group may be the most appropriate.

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press

In some investigations with an individually matched design, the population controls are the neighbours, friends or relatives of the cases (see, for example, Amaducci *et al*, 1986).

(3) Volunteer controls

Healthy volunteers are most commonly used in studies where a blood, imaging or other test is being investigated in psychiatric disorders. In this situation, it is difficult to assess the potential biases which may arise and, if volunteers are used, the investigators must make every attempt to describe who they were, why they volunteered, and what the response rate was.

In a recent study exploring positron emission tomography (PET) scan abnormalities in obsessive-compulsive disorder, no details were given on who the controls were and whether they differed markedly from the study population (Baxter *et al*, 1987). The general point made above, that controls should be included in the sampling frame for cases if they become diseased, tends to outlaw the use of volunteer controls in most psychiatric centres.

(4) Case-control studies and surveys

The widely acknowledged problem of selection bias in studying affective disorder has led to the adoption of community surveys to investigate the disorder. This ensures that both the cases and controls come from the same population and reduces the selection bias. Fortunately, many psychiatric conditions are common enough to adopt this approach, but in studying rarer disorders such as schizophrenia or severe depressions, surveys share the same disadvantage as cohort studies in requiring large numbers of subjects to detect small numbers of cases.

In studying common minor psychiatric disorders, one approach may be to screen a population and study a smaller sample of likely cases and controls in more depth. This has been called a two-stage screening procedure (Williams *et al*, 1980; Bebbington *et al*, 1981).

Examples of selection bias

Brown & Harris (1978) interviewed a sample of in-patients and out-patients in addition to a general population sample. The results (Table IV) illustrate that the association between having three or more children under 14 years at home and the onset of depression was only apparent when the general population cases were compared with the community controls (odds ratio = 3.77) and not when patient cases were compared with them. This is an example of selection into treatment. The presence of young children may increase the risk of depression, and also reduce the risk of receiving psychiatric care.

TABLE IV

An example of selection into treatment (from Brown & Harris (1978)): presence of more than 3 children (< 14 years) in patient and general population cases of depression as well as in general population controls

		Patient cases	*General population cases*	*General population controls*
>3 children	yes	9	9	30
(< 14 years)	no	105	28	352

Odds ratios comparing cases with general population controls: for patient cases = 1.01 (95% CI: 0.47, 2.19); for general population cases = 3.77 (95% CI: 1.63, 8.72).

The study by McGuffin *et al* (1988) may also be subject to selection bias. They noted an increase in the incidence of life events in the relatives of depressed hospital patients compared with the results of a population survey. Their results though suggest that the relatives were not very representative, because there was a marked preponderance of women (145:99). In addition, many relatives were not living in the Camberwell area when the study was conducted and so their comparison with Camberwell community data is less than ideal. Furthermore, they began by choosing relatives of depressed hospital patients who are unrepresentative of depressed people in the community. A case-control study could try to minimise the possible selection bias by choosing an appropriate control group.

Information bias

Psychiatric researchers have long been concerned with measurement because of the serious distortions that can result from inaccurate or biased assessment.

A distinction is made between random and systematic error. Random error will tend to bias the relative risk towards unity. A more serious problem arises from a systematic error: when there is a difference between the cases and controls in the gathering of data on disease or exposure. Such systematic error could lead to either an *increase* or *decrease* in the relative risk.

It is convenient to divide information bias into that derived from the subject and that from the observer although more exhaustive classifications exist (see Sackett, 1979).

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press

Subject bias

Most case-control studies ask about exposure retrospectively and this leads to a serious source of bias, often termed 'recall bias': that subjects' recall of exposure may be influenced by their disease status.

One of the best known examples of recall bias is Stott's (1958) finding that mothers of children with Down's syndrome more frequently reported an emotional stressor during pregnancy than women who delivered a healthy baby. This is now thought to be due to the mothers' attempt to explain bearing a handicapped child – an 'effort after meaning'.

Recall bias is of importance in life-events research. A depressed subject may remember aversive life events more readily than someone without depression (Lloyd & Lishman, 1976). This would increase the apparent association between life events and depression.

There are several strategies for minimising recall bias and these should be considered at the design stage.

(a) Using a structured questionnaire and standardising the criteria for exposure is a prerequisite for reducing information bias, including recall bias. Relying upon recognition rather than recall by supplying the subject with a comprehensive series of prompts should also minimise this bias. This method is now routinely adopted in life-events research.

(b) Since recall bias is less likely to occur if the illness is of recent onset, the study of incident rather than prevalent cases is favoured.

(c) Control subjects who are themselves suffering from an illness may be equally motivated to search their memories with regard to exposure. A more appropriate comparison in the Down's syndrome study above would have been with the mothers of babies with a different handicap of known cause rather than the mothers of healthy babies.

(d) The validity of self-reported exposure data can be tested, for example by interviewing a close relative or consulting medical records.

(e) Raphael (1987) suggests that if other intuitively plausible exposures, which are known not to be risk factors, are not differentially recalled by the study groups, this lends weight to a finding of association with the hypothesised exposure. In addition, subjects could be asked about their own causal hypothesis, to see if this corresponds with that of the experimenter.

Social network (Henderson *et al*, 1981), personality measures (Katz & McGuffin, 1987), early aversive experiences (Wolkind & Coleman, 1983) and marital discord are probably influenced by current mental state. Therefore, case-control methodology and cross-sectional surveys may lead to misleading results. One potential approach is to design a case-control study within a cohort. For instance, Brown *et al* (1987) followed a group of women who were at increased risk of becoming depressed and collected data liable to be influenced by mental state at the beginning of the follow-up period. They then compared those who became depressed with those who did not using a case-control design. A similar design is often adopted (e.g. Kumar & Robson, 1984) in attempting to study risk factors for post-natal depression.

Observer bias

When an interviewer knows the subject is a case there may be biased assessment of the exposure, for instance by more careful probing. Similarly, knowledge of exposure, for instance a family member with mental illness, may influence diagnosis in doubtful cases. For this reason, epidemiologists have long argued against an over-reliance on unstructured clinical judgements in defining disease and exposure. The concept of 'blind' assessment in clinical trials (Hill, 1965) arose from the observation that knowledge of the treatment condition can seriously bias the results of a clinical assessment. Similar standards must also pertain to work studying associations in observational studies and so 'blind' assessments should be made whenever possible. In practice, this is often difficult because the author of the research proposal also acts as interviewer or because it is hard to sustain the interest and morale of the research team if they are unaware of the study purpose.

Epidemiologists who investigate the aetiology of physical diseases prefer to use non-medically trained interviewers as they are more likely to follow the questionnaire exactly and not to use their own judgement in deciding on exposure status. Another possibility is to tape-record interviews which can then be co-rated, preferably by someone who can be kept unaware of whether the subject is a case or a control. In research using imaging or laboratory techniques it is relatively easy for assessments to be performed by a rater who is unaware of whether the subject is a case or a control.

Observer bias is eliminated by self-report questionnaires and computerised assessments (Lewis *et al*, 1988). On the other hand, there is concern that subjects may not fully understand self-report questionnaires, and that those who read poorly or are illiterate may be unable to respond. Other sources

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press

of bias, for instance when the subject's responses are influenced by disease status (see subject bias), will remain when using self-report measures. Nevertheless, the reliance on clinical judgement in defining cases of psychiatric disorder in most psychiatric interviews is an unresolved problem as it may increase observer bias.

Measuring bias

If it is possible to measure a bias then it can be treated as a confounder. Unquantifiable biases are a major problem because it becomes difficult to interpret the results of a study. A bias may have a trivial effect or alternatively lead to a major distortion. The only solution to this difficulty is to attempt to measure bias, although "efforts to compensate for it [bias] in the analysis must be viewed with suspicion" (Rose & Barker, 1978).

An example of possible information bias

Goldberg & Morrison (1963) compared the social class of the parents of schizophrenic patients with that of the general population. As discussed above, the general population is probably an appropriate sample to compare with schizophrenic first admissions. However, they ascertained parental social class from the birth certificates of the schizophrenics and compared this with census data on the general population. Social class was therefore collected in different ways in cases and 'controls'. This can result in a well documented bias as social class tends to be 'higher' when recorded on birth certificates than when recorded by census (Leete & Fox, 1977). This suggests that a traditional case-control design would be more appropriate and would enable the social class of both cases and controls to be ascertained in the same way.

Conclusion

It should be clear that almost all the problems can (or should) be foreseen in the design of the study and the main intellectual challenge of conducting a study comes before the data are collected; in choosing the right control group, deciding on potential confounders and their appropriate measurement and assessing the importance of biases. It is therefore invaluable to consult colleagues at this stage and approach an epidemiologist and statistician, including those who might work outside the field of psychiatry.

Many authors feel reluctant to discuss the weaknesses of their own experimental design. This understandable but unfortunate failing hinders scientific progress. Only by an open discussion of weaknesses as well as strengths can new strategies be developed to overcome methodological problems.

Case-control studies provide both a challenge and an opportunity to the researcher. This has been well put by Breslow & Day (1980): "as a form of research the case-control study continues to offer a paradox: compared with other designs it is a rapid and efficient way to evaluate a hypothesis. On the other hand, despite its practicality, the case-control study is not simplistic and it cannot be done well without considerable planning. Indeed, the case-control study is perhaps the most challenging to design and conduct in such a way that bias is avoided. Our limited understanding of this difficult study design and its many subtleties should serve as a warning – these studies must be designed and analysed carefully with a thorough appreciation of their difficulties."

Appendix

Simple calculations in unmatched case-control studies

The odds ratio, R, and its confidence intervals will be calculated. Most authors will also carry out a χ^2 test of statistical significance. Table V shows the results of an unmatched case-control study in algebraic form. The odds ratio, which conforms to the relative risk as long as the condition is relatively rare, is calculated as follows:

Odds of being a case in exposed group = a/b
Odds of being a case in unexposed group = c/d
Odds ratio (exposed/unexposed) = (a/b)/(c/d) = ad/bc

There are a number of ways of calculating the confidence intervals; Woolf's method is given here.

The logarithm of the odds ratio (ln R) is approximately normally distributed. 95% confidence intervals for the odds ratio can then be calculated as follows:

$$\ln R \pm 1.96 \times \sqrt{\{1/a + 1/b + 1/c + 1/d\}}$$

Antilogs of the resulting limits give upper and lower 95% confidence limits for the odds ratio. When any cell in the table equals zero, 0.5 should be added to each cell before the variance of the logarithm of the odds ratio is calculated.

Matched case-control studies

Different methods of analysis are needed for matched case-control studies and conditional logistic regression is needed for multivariate analysis.

TABLE V

General 2×2 table showing the frequency of exposure in an unmatched sample of n_1 cases and n_2 controls

	Cases	*Controls*	*Total*
Exposed	a	b	m_1
Not exposed	c	d	m_2
Total	n_1	n_2	n

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press

Further accounts

A number of good textbooks exist that discuss case-control studies in more depth. Schlesselman (1982) deals exclusively with case-control studies while Hennekens & Buring (1987) and the more difficult Rothman (1986) also discuss other areas of epidemiology. Breslow & Day (1980) have written a detailed and authoritative book on the analysis of case-control studies.

Acknowledgements

We would like to thank Professor Geoffrey Rose, Professor Michael Shepherd and the anonymous referees for comments. We also thank Professor Rose and his colleagues at the London School of Hygiene and Tropical Medicine for their teaching. The General Practice Research Unit is funded by the Department of Health and was under the direction of Professor Michael Shepherd. GL is funded by the Health Promotion Research Trust. AJP is a Wellcome Research Fellow in Clinical Epidemiology.

References

AMADUCCI, L. A., FRATIGLIONI, L., ROCCA, W., *et al* (1986) Risk factors for clinically diagnosed Alzheimer's disease: a case-control study of an Italian population. *Neurology*, **36**, 922–931.

ANDREASSEN, S., ALLEBECK, P., ENGSTROM, A., *et al* (1987) Cannabis and schizophrenia: a longitudinal study of Swedish conscripts. *Lancet*, *ii*, 1483–1485.

BAXTER, L. R., PHELPS, M. E., MAZZIOTA, J. C., *et al* (1987) Local cerebral glucose metabolic rates in obsessive-compulsive disorder. *Archives of General Psychiatry*, **44**, 211–218.

BEBBINGTON, P., HURRY, J., TENNANT, C., *et al* (1981) Epidemiology of mental disorders in Camberwell. *Psychological Medicine*, **11**, 561–579.

——, STURT, E., TENNANT, C., *et al* (1984) Misfortune and resilience: a community study of women. *Psychological Medicine*, **14**, 347–363.

BRESLOW, N. E. & DAY, N. E. (1980) *Statistical Methods in Cancer Research: Vol. 1, The Analysis of Case-Control Studies*. Oxford: Oxford University Press.

BROWN, G. W. & HARRIS, T. (1978) *Social Origins of Depression*. London: Tavistock.

——, BIFULCO, A. & HARRIS, T. O. (1987) Life events, vulnerability and the onset of depression: some refinements. *British Journal of Psychiatry*, **150**, 30–42.

CARROLL, B. J., FEINBERG, M., GREDEN, J. F., *et al* (1981) A specific laboratory test for the diagnosis of melancholia. *Archives of General Psychiatry*, **38**, 15–22.

CORYELL, W. & TSUANG, M. T. (1982) DSM-III schizophreniform disorder. *Archives of General Psychiatry*, **39**, 66–69.

CRAIG, T. K. & BROWN, G. W. (1984) Goal frustrating aspects of life events, stress and aetiology of gastro-intestinal disorder. *Journal of Psychosomatic Research*, **28**, 411–421.

DOLL, R. & HILL, A. B. (1952) A study of the aetiology of carcinoma of the lung. *British Medical Journal*, *ii*, 1271–1286.

DUNN, G. (1981) The role of linear models in psychiatric epidemiology. *Psychological Medicine*, **11**, 179–184.

EATON, W. W. (1985) Epidemiology of schizophrenia. *Epidemiologic Reviews*, **7**, 105–126.

GARDNER, M. J. & ALTMAN, D. G. (1986) Confidence intervals rather than P values: estimation rather than hypothesis testing. *British Medical Journal*, **292**, 746–750.

GOLDBERG, D. & HUXLEY, P. (1980) *Mental Illness in the Community*. London: Tavistock.

GOLDBERG, E. M. & MORRISON, S. L. (1963) Schizophrenia and social class. *British Journal of Psychiatry*, **109**, 785–802.

HENDERSON, A. S., BYRNE, D. G. & DUNCAN-JONES, P. (1981) *Neurosis and the Social Environment*. Sydney: Academic Press.

HENNEKENS, C. H. & BURING, J. E. (1987) *Epidemiology in Medicine*. Boston: Little Brown.

HILL, A. B. (1965) The environment and disease: association or causation? *Journal of the Royal Society of Medicine*, **58**, 295–300.

JENKINS, R. (1985) Sex differences in minor psychiatric disorder. *Psychological Medicine*, monograph no. 7.

KATZ, R. & MCGUFFIN, P. (1987) Neuroticism in familial depression. *Psychological Medicine*, **17**, 155–161.

KUMAR, R. & ROBSON, K. M. (1984) A prospective study of emotional disorders in childbearing women. *British Journal of Psychiatry*, **144**, 35–47.

LEETE, R. & FOX, J. (1977) Registrar General's social classes: origins and uses. *Population Trends*, **8**, 1–7.

LEWIS, G. H., PELOSI, A. H., GLOVER, E., *et al* (1988) The development of a computerised assessment for minor psychiatric disorder. *Psychological Medicine*, **18**, 737–745.

LEWIS, S.W. & MURRAY, R. M. (1987) Obstetric complications, neurodevelopmental deviance and risk of schizophrenia. *Journal of Psychiatric Research*, **21**, 413–421.

LLOYD, G. & LISHMAN, A. (1976) The effect of depression on speed of recall of pleasant and unpleasant experiences. *Psychological Medicine*, **5**, 173–180.

MANTEL, N. & HAENZEL, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, **22**, 719–748.

MCGUFFIN, P., KATZ, R., ALDRICH, J., *et al* (1988) The Camberwell collaborative depression study. II: Investigation of family members. *British Journal of Psychiatry*, **152**, 766–774.

MUKHERJEE, S., ROSEN, A. M., CARACCI, G., *et al* (1986) Persistent tardive dyskinesia in bipolar patients. *Archives of General Psychiatry*, **43**, 342–346.

MULLEN, P. E., LINSELL, C. R. & PARKER, D. (1986) Influence of sleep disruption and calorie restriction on biological markers for depression. *Lancet*, *ii*, 1051–1054.

RAPHAEL, K. (1987) Recall bias: a proposal for assessment for control. *International Journal of Epidemiology*, **16**, 167–170.

ROSE, G. & BARKER, D. J. P. (1978) Observer variation. *British Medical Journal*, *ii*, 1006–1007.

ROTHMAN, K. (1986) *Modern Epidemiology*. Boston: Little Brown.

SACKETT, D. L. (1979) Bias in analytic research. *Journal of Chronic Diseases*, **32**, 51–63.

SCHLESSELMAN, J. J. (1982) *Case Control Studies*. New York: Oxford University Press.

SHEPHERD, M., COOPER, B., BROWN, A. C., *et al* (1966) *Psychiatric Illness in General Practice*. Oxford: Oxford University Press.

SMITH, G. N., IACONO, W. G., MOREAU, M., *et al* (1988) Choice of comparison group and findings of computerised tomography in schizophrenia. *British Journal of Psychiatry*, **153**, 667–674.

SMITH, P. G., RODRIGUES, L. C. & FINE, P. E. M. (1984) Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. *American Journal of Epidemiology*, **13**, 87–93.

STOTT, D. H. (1958) Some psychosomatic aspects of causality in reproduction. *Journal of Psychosomatic Research*, **3**, 42–55.

WARR, P. (1987) *Work, Unemployment and Mental Health*. Oxford: Oxford University Press.

WILLIAMS, P., TARNOPOLSKY, A. & HAND, D. (1980) Case definition and case identification in psychiatric epidemiology: review and assessment. *Psychological Medicine*, **10**, 101–114.

WOLKIND, S. & COLEMAN, E. Z. (1983) Adult psychiatric disorder and childhood experiences: the validity of retrospective data. *British Journal of Psychiatry*, **143**, 188–191.

*Glyn Lewis, MSc, MRCPsych, *Lecturer, General Practice Research Unit, Institute of Psychiatry*; Anthony J. Pelosi, MSc, MRCP, MRCPsych, *Research Fellow in Clinical Epidemiology, Institute of Psychiatry*

*Correspondence: *Institute of Psychiatry, DeCrespigny Park, London SE5 8AF*

https://doi.org/10.1192/bjp.157.2.197 Published online by Cambridge University Press