Bias and causal associations in observational research

David A Grimes, Kenneth F Schulz

Readers of medical literature need to consider two types of validity, internal and external. Internal validity means that the study measured what it set out to; external validity is the ability to generalise from the study to the reader’s patients. With respect to internal validity, selection bias, information bias, and confounding are present to some degree in all observational research. Selection bias stems from an absence of comparability between groups being studied. Information bias results from incorrect determination of exposure, outcome, or both. The effect of information bias depends on its type. If information is gathered differently for one group than for another, bias results. By contrast, non-differential misclassification tends to obscure real differences. Confounding is a mixing or blurring of effects: a researcher attempts to relate an exposure to an outcome but actually measures the effect of a third factor (the confounding variable). Confounding can be controlled in several ways: restriction, matching, stratification, and more sophisticated multivariate techniques. If a reader cannot explain away study results on the basis of selection, information, or confounding bias, then chance might be another explanation. Chance should be examined last, however, since these biases can account for highly significant, though bogus results. Differentiation between spurious, indirect, and causal associations can be difficult. Criteria such as temporal sequence, strength and consistency of an association, and evidence of a dose-response effect lend support to a causal link.

Clinicians face two important questions as they read medical research: is the report believable, and, if so, is it relevant to my practice? Uncritical acceptance of published research has led to serious errors and squandered resources. Here, we will frame these two questions in terms of study validity, describe a simple checklist for readers, and offer some criteria by which to judge reported associations.

Internal and external validity

Analogous to a laboratory test, a study should have internal validity—ie, the ability to measure what it sets out to measure. The inference from participants in a study should be accurate. In other words, a research study should avoid bias or systematic error. Internal validity is the sine qua non of clinical research; extrapolation of invalid results to the broader population is not only worthless but potentially dangerous.

A second important concern is external validity; can results from study participants be extrapolated to the reader’s patients? Since a total enumeration or census approach to medical research is usually impossible, the customary tactic is to choose a sample, study it, and, hopefully, extrapolate the result to one’s practice. Gauging external validity is necessarily more subjective than is assessment of internal validity.

Internal and external validity entail important trade-offs. For example, randomised controlled trials are more likely than observational studies to be free of bias, but, because they usually enrol selected participants, external validity can suffer. This problem of unsuitable participants is also termed distorted assembly. Participants in randomised controlled trials tend to be different (including being healthier*) from those who choose not to take part, a function of the restricted entry criteria. The filtering process for admission to randomised trials might, therefore, result in “a type of hothouse flower, which cannot bloom or be successfully removed beyond its special greenery.”

Bias

Bias undermines the internal validity of research. Unlike the conventional meaning of bias—ie, prejudice—bias in research denotes deviation from the truth. All observational studies (and, regrettably, many badly done randomised controlled trials) have built-in bias; the challenge for investigators, editors, and readers is to ferret these out and judge how they might have affected results. A simple checklist, such as that shown in panel 1, can be helpful.

Several taxonomies exist for classification of biases in clinical research. Sackett’s landmark compilation, for example, included 35 different biases. By contrast Feinstein* consolidated biases into four categories that arise sequentially during research: susceptibility, performance, detection, and transfer. Susceptibility bias refers to differences in baseline characteristics, performance bias to different proficiencies of treatment, detection bias to different measurement of outcomes, and transfer bias to differential losses to follow-up. Another approach which is often used, is to group all biases into three general categories: selection, information, and confounding. The leitmotif for all three is “different”. Something “different” distorts the planned comparison.

Selection bias

Are the groups similar in all important respects?

Selection bias stems from an absence of comparability between groups being studied. For example, in a cohort study, the exposed and unexposed groups differ in some important respect aside from the exposure. Membership bias is a type of selection bias: people who choose to be members of a group—eg, joggers—might differ in important respects from others. For instance, both cohort and case-control studies initially suggested that jogging after myocardial infarction prevented repeat

Lancet 2002; 359: 248–52

Family Health International, PO Box 13950, Research Triangle Park, NC 27709, USA (D A Grimes MD, K F Schulz PhD)

Correspondence to: Dr David A Grimes (e-mail: dgrimes@fhi.org)
infarction. However, a randomised controlled trial failed to confirm this benefit. Those who chose to exercise might have differed in other important ways from those who did not exercise, such as diet, smoking, and presence of angina.

In case-control studies, selection bias implies that cases and controls differ importantly aside from the disease in question. Two types of selection bias have earned eponyms: Berkson and Neyman bias. Also known as an admission-rate bias, Berkson bias (or paradox) results from differential rates of hospital admission for cases and controls. Berkson initially thought that this phenomenon was due to presence of a simultaneous disease. Alternatively, knowledge of the exposure of interest might lead to an increased rate of admission to hospital. For example, doctors who care for women with subclinical cancer, leading to a spurious increase in the odds ratio. In Swedish and Danish cohort studies the disease among those exposed, for example, preferentially searching for infection by HIV-1 in intravenous drug users. Conversely, the presence of a disease might prompt a search for the putative exposure of interest. Another type of bias is family history bias, in which medical information flows differently to affected and unaffected family members, as has been shown for rheumatoid arthritis.

Information bias can arise in many ways. Some use the term ascertainment to describe gathering information in different ways. For example, an investigator might gather information about an exposure at bedside for a case but by telephone from a community control. Diagnostic suspicion bias implies that knowledge of a putative cause of disease might launch a more intensive search for the disease among those exposed, for example, preferentially interviewing abortion history of prior abortions, obtained by personal interview, against centralised medical records, they documented systematic underreporting of abortions among controls (but not among cases) that accounted for a spurious association.

Panel 1: What to look for in observational studies

Is selection bias present?
In a cohort study, are participants in the exposed and unexposed groups similar in all important respects except for the exposure?
In a case-control study, are cases and controls similar in all important respects except for the disease in question?

Is information bias present?
In a cohort study, is information about exposure gathered in the same way for those exposed and unexposed?
In a case-control study, is information about exposure gathered in the same way for cases and controls?

Is confounding present?
Could the results be accounted for by the presence of a factor—eg, age, smoking, sexual behaviour, diet—associated with both the exposure and the outcome but not directly involved in the causal pathway?

If the results cannot be explained by these three biases, could they be the result of chance?
What are the relative risk or odds ratio and 95% CI?
Is the difference statistically significant, and, if so, is the study design adequate to find a clinically important difference?

If the results still cannot be explained away, then (and only then) might the findings be real and worthy of note.

therapy might cause symptomless endometrial cancer patients to bleed, resulting in initiation of diagnostic tests. In this instance, the exposure unmasked the subclinical cancer, leading to a spurious increase in the odds ratio. In observational studies, non-respondents are different from respondents. Cigarette smokers are a case in point: smokers are less likely to return questionnaires than are non-smokers or pipe and cigar smokers.

Information bias

Has information been gathered in the same way?
Information bias, also known as observation, classification, or measurement bias, results from incorrect determination of exposure or outcome, or both. In a cohort study or randomised controlled trial, information about outcomes should be obtained the same way for those exposed and unexposed. In a case-control study, information about exposure should be gathered in the same way for cases and controls.

Information bias can arise in many ways. Some use the term ascertainment to describe gathering information in different ways. For example, an investigator might gather information about an exposure at bedside for a case but by telephone from a community control. Diagnostic suspicion bias implies that knowledge of a putative cause of disease might launch a more intensive search for the disease among those exposed, for example, preferentially searching for infection by HIV-1 in intravenous drug users. Conversely, the presence of a disease might prompt a search for the putative exposure of interest. Another type of bias is family history bias, in which medical information flows differently to affected and unaffected family members, as has been shown for rheumatoid arthritis.

To minimise information bias, detail about exposures in case-control studies should be gathered by people who are unaware of whether the respondent is a case or a control. Similarly, in a cohort study with subjective outcomes, the observer should be unaware of the exposure status of each participant.

In case-control studies that rely on memory of remote exposures, recall bias is pervasive. Cases tend to search their memories to identify what might have caused their disease; healthy controls have no such motivation. Thus, better recall among cases is common. For example, the putative association between induced abortion and subsequent development of breast cancer has emerged as a hot medical and political issue. Many case-control studies have reported an increase in cancer risk after abortion. However, when investigators compared histories of prior abortions, obtained by personal interview, against centralised medical records, they documented systematic underreporting of abortions among controls (but not among cases) that accounted for a spurious association.

In Swedish and Danish cohort studies, free from recall bias, induced abortion has had either a protective effect or no effect on risk of breast cancer.
Confounding

Is an extraneous factor blurring the effect? Confounding is a mixing or blurring of effects. A researcher attempts to relate an exposure to an outcome, but actually measures the effect of a third factor, termed a confounding variable. A confounding variable is associated with the exposure and it affects the outcome, but it is not an intermediate link in the chain of causation between exposure and outcome.27,28 More simply, confounding is a methodological fly in the ointment. Confounding is often easier to understand from examples than from definitions.

Oral contraceptives and myocardial infarction, and smoking

Early studies of the safety of oral contraceptives reported a pronounced increased risk of myocardial infarction. This association later proved to be spurious, because of the high proportion of cigarette smokers among users of birth control pills.29–31 Here, cigarette smoking confounded the relation between oral contraceptives and infarction. Women who chose to use birth control pills also chose, in large numbers, to smoke cigarettes, and cigarettes, in turn, increased the risk of myocardial infarction. Although investigators thought they were measuring an effect of birth control pills, they were in fact measuring the hidden effect of smoking among pill users.

IUD insertion and salpingitis, and exposure to sexually transmitted disease

Results of a large case-control study of IUDs indicated a significant increase in salpingitis soon after insertion.10 However, among married or cohabiting women with only one reported sex partner in the past 6 months, no significant increase in risk was evident.11 In the study, exposure to sexually transmitted diseases apparently confounded the association. Even among women at low risk of salpingitis, frequent coitus might increase risk of infection,2 and few studies have controlled for this variable.

Oral contraceptives and cervical cancer, and smoking

Reported associations between oral contraceptives and squamous cervical cancer3 might be due to unsuspected confounding by cigarette smoking and human papillomavirus infection.32 Control of confounding is inevitably limited by our meagre understanding of human biology; unsuspected confounding factors evade control in observational studies.33

Control for confounding

When selection bias or information bias exist in a study, irreparable damage results. Internal validity is doomed. By contrast, when confounding is present, this bias can be corrected, provided that confounding was anticipated and the requisite information gathered. Confounding can be controlled for before or after a study is done. The purpose of these approaches is to achieve homogeneity between study groups.

Restriction

The simplest approach is restriction (also called exclusion or specification).29 For example, if cigarette smoking is suspected to be a confounding factor, a study can enrol only non-smokers. Although this tactic avoids confounding, it also hinders recruitment (and thus power) and precludes extrapolation to smokers. Restriction might increase the internal validity of a study at the cost of poorer external validity.

Matching

Another way to control for confounding is pairwise matching. In a case-control study in which smoking is deemed a confounding factor, cases and controls can be matched by smoking status. For each case who smokes, a control who smokes is found. This approach, although often used by investigators, has two drawbacks. If matching is done on several potential confounding factors, the recruitment process can be cumbersome, and, by definition, one cannot examine the effect of a matched variable.28

Stratification

Investigators can also control for confounding after a study has been completed. One approach is stratification. Stratification can be considered a form of post hoc restriction, done during the analysis rather than during the accrual phase of a study. For example, results can be stratified by levels of the confounding factor. In the smoking example, results are calculated separately for smokers and non-smokers to see if the same effect arises independent of smoking. The Mantel-Haenszel procedure34 combines the various strata into a summary statistic that describes the effect. The strata are weighted inversely to their variance—ie, strata with larger numbers count more than those with smaller numbers. If the Mantel-Haenszel adjusted effect differs substantially from the crude effect, then confounding is deemed present. In this instance, the adjusted estimate of effect is considered the better estimate to use.

Confounding is not always intuitive, as shown by the fictitious example in the figure. In this hypothetical

Example of confounding in a hypothetical cohort study of intrauterine device use and salpingitis

When the crude relative risk is controlled for the confounding effect of number of sexual partners, the raised risk disappears.
For personal use. Only reproduce with permission from The Lancet Publishing Group.
smoke on the bronchial epithelium of animals is coherent with an increased risk of cancer in human beings. Finally, experimental evidence is seldom available, and reasoning by analogy has sometimes caused harm. Since thalidomide can cause birth defects, for instance, some lawyers (successfully) argued by analogy that Bendectin (an antecmetic widely used for nausea and vomiting in pregnancy) could also cause birth defects, despite evidence to the contrary.  

Conclusion
Studies need to have both internal and external validity: the results should be both correct and capable of extrapolation to the population. A simple checklist for bias (selection, information, and confounding) then chance can help readers decipher research reports. When a statistical association appears in research, guidelines for judgment of associations can help a reader decide whether the association is bogus, indirect, or real.

We thank Willard Cates and David L Sackett for their helpful comments on an earlier version of this report. Much of this material stems from our 15 years of teaching the Berlex Foundation Faculty Development Course.

References