Epidemiology series

An overview of clinical research: the lay of the land

David A Grimes, Kenneth F Schulz

Many clinicians report that they cannot read the medical literature critically. To address this difficulty, we provide a primer of clinical research for clinicians and researchers alike. Clinical research falls into two general categories: experimental and observational, based on whether the investigator assigns the exposures or not. Experimental trials can also be subdivided into two: randomised and non-randomised. Observational studies can be either analytical or descriptive. Analytical studies feature a comparison (control) group, whereas descriptive studies do not. Within analytical studies, cohort studies track people forward in time from exposure to outcome. By contrast, case-control studies work in reverse, tracing back from outcome to exposure. Cross-sectional studies are like a snapshot, which measures both exposure and outcome at one time point. Descriptive studies, such as case-series reports, do not have a comparison group. Thus, in this type of study, investigators cannot examine associations, a fact often forgotten or ignored. Measures of association, such as relative risk or odds ratio, are the preferred way of expressing results of dichotomous outcomes—eg, sick versus healthy. Confidence intervals around these measures indicate the precision of these results. Measures of association with confidence intervals reveal the strength, direction, and a plausible range of an effect as well as the likelihood of chance occurrence. By contrast, p values address only chance. Testing null hypotheses at a p value of 0·05 has no basis in medicine and should be discouraged.

Clinicians today are in a bind. Increasing demands on their time are squeezing out opportunities to stay abreast of the literature, much less read it critically. Results of several studies indicate an inverse relation between knowledge of contemporary care and time since graduation from medical school. In many jurisdictions, attendance at a specified number of hours of continuing medical education courses is mandatory to maintain a licence to practise. However, the failure of these courses to improve patient care emphasises the importance of self-directed learning through reading. Many clinicians in practice, though, report that they feel unqualified to read the medical literature critically. Scientific illiteracy is a problem, though, the study design must be known. As in biology, anatomy dictates physiology. The anatomy of a study determines what it can and cannot do. A difficulty that many clinicians report is that authors sometimes do not report the methodological elements of trials that minimise bias. Finally, because trials are so important, clinicians might be more likely to act on their results than on those of observational studies; hence, investigators need to ensure that trials are done and reported well. Here, we provide a brief overview of research designs and discuss some of the common measures used.

A taxonomy of clinical research

Analogue to biological taxonomy, a simple hierarchy can be used to categorise most studies (panel). To do so, however, the study design must be known. As in biology, anatomy dictates physiology. The anatomy of a study determines what it can and cannot do. A difficulty that readers encounter is that authors sometimes do not report the methodological elements of trials that minimise bias. Finally, because trials are so important, clinicians might be more likely to act on their results than on those of observational studies; hence, investigators need to ensure that trials are done and reported well. Here, we provide a brief overview of research designs and discuss some of the common measures used.

Rating clinical evidence

Assessment system of the US Preventive Services Task Force

Quality of evidence

I Evidence from at least one properly designed randomised controlled trial.
II-1 Evidence obtained from well-designed controlled trials without randomisation.
II-2 Evidence from well-designed cohort or case-control studies, preferably from more than one centre or research group.
II-3 Evidence from multiple time series with or without the intervention. Important results in uncontrolled experiments (such as the introduction of penicillin treatment in the 1940s) could also be considered as this type of evidence.
III Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.

Strength of recommendations

A Good evidence to support the intervention.
B Fair evidence to support the intervention.
C Insufficient evidence to recommend for or against the intervention, but recommendation might be made on other grounds.
D Fair evidence against the intervention.
E Good evidence against the intervention.

Lancet 2002; 359: 57–61

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)
versus restricted access to electronic fetal monitoring for the latter would be a trial alternating months of liberal was used, such as alternate assignment. An example of those involved) or whether some other allocation scheme (with concealment of the upcoming assignment from the investigators assigned the exposures—eg, treatments—or whether they observed usual clinical practice. For experimental studies, one needs to distinguish whether the exposures were assigned by a truly random technique (with concealment of the upcoming assignment from those involved) or whether some other allocation scheme was used, such as alternate assignment. An example of the latter would be a trial alternating months of liberal versus restricted access to electronic fetal monitoring for women in labour.

With observational studies, which dominate the literature, the next step is to ascertain whether the study has a comparison or control group. If so, the study is termed analytical. If not, it is a descriptive study (figure 1). If the study is analytical, the temporal direction of the trial needs to be identified. If the study determines both exposures and outcomes at one time point, it is termed cross-sectional. An example would be measurement of serum cholesterol of men admitted to a hospital with myocardial infarction versus that of their next-door neighbour. This type of study provides a snapshot of the population of sick and well at one time point.

If the study begins with an exposure—eg, oral contraceptive use—and follows women for a few years to measure outcomes—eg, ovarian cancer—then it is deemed a cohort study. Cohort studies can be either concurrent or non-concurrent. By contrast, if the analytical study begins with an outcome—eg, ovarian cancer—and looks back in time for an exposure, such as use of oral contraceptives, then the study is a case-control study.

Studies without comparison groups are called descriptive studies. At the bottom of the research hierarchy is the case report. When more than one patient is described, it becomes a case-series report.

**What studies can and cannot do**

*Is the study design appropriate for the question?*

Starting at the bottom of the research hierarchy, descriptive studies are often the first foray into a new area of medicine. Investigators do descriptive studies to describe the frequency, natural history, and possible determinants of a condition. The results of these studies show how many people develop a disease or condition over time, describe the characteristics of the disease and those affected, and generate hypotheses about the cause of the disease. These hypotheses can be assessed through more rigorous research, such as analytical studies or randomised controlled trials. An example of a descriptive study would be the early reports of Legionnaire’s disease and toxic-shock syndrome. An important caveat (often forgotten or intentionally ignored) is that descriptive studies, which do not have a comparison group, do not allow assessment of associations. Only comparative studies (both analytical and experimental) enable assessment of possible causal associations.

**Cross-sectional study: a snapshot in time**

Sometimes termed a frequency survey or a prevalence study, cross-sectional studies are done to examine the presence or absence of disease and the prevalence or absence of an exposure at a particular time. Thus, prevalence, not incidence, is the focus. Since both outcome and exposure are ascertained at the same time (figure 2), the temporal relation between the two might be unclear. For example, assume that a cross-sectional study finds obesity to be more common among women with than without arthritis. Did the extra weight load on joints lead to arthritis, or did women with arthritis become involuntarily inactive and then obese? This type of question is unanswerable in a cross-sectional study.

**Cohort study: looking forward in time**

Cohort studies proceed in a logical sequence: from exposure to outcome (figure 2). Hence, this type of research is easier to understand than case-control studies. Investigators identify a group with an exposure of interest and another group or groups without the exposure. The investigators then follow the exposed and unexposed groups forward in time to determine outcomes. If the exposed group develops a higher incidence of the outcome than the unexposed, then the exposure is associated with an increased risk of the outcome.

The cohort study has important strengths and weaknesses. Because exposure is identified at the outset, one can assume that the exposure preceded the outcome. Recall bias is less of a concern than in the case-control study. The cohort study enables calculation of true incidence rates, relative risks, and attributable risks. However, for the study of rare events or events that take years to develop, this type of research design can be slow to yield results and thus prohibitively expensive. Nonetheless, several famous, large cohort studies continue to provide important information.
Outbreaks of food-borne diseases are a prototype for case-control studies work backwards. Because thinking in this direction is not intuitive for clinicians, case-control studies are widely misunderstood. Starting with an outcome, such as disease, this type of study looks backward in time for exposures that might have caused the outcome. As shown in figure 2, investigators define a group with an outcome (for example, ovarian cancer) and a group without the outcome (controls). Then, through chart reviews, interviews, or other means, the investigators ascertain the prevalence (or amount) of exposure to a risk factor—eg, oral contraceptives, ovulation-induction drugs—in both groups. If the prevalence of the exposure is higher among cases than among controls, then the exposure is associated with an increased risk of the outcome.

Randomised controlled trial: gold standard
The randomised controlled trial is the only known way to avoid selection and confounding biases in clinical research. This design approximates the controlled experiment of basic science. It resembles the cohort study in several respects, with the important exception of randomisation of participants to exposures (figure 2). The hallmark of randomised controlled trials is assignment of participants to exposures purely by the play of chance. Randomised controlled trials reduce the likelihood of bias in determination of outcomes. When properly implemented, random allocation precludes selection bias. Trials feature uniform diagnostic criteria for outcomes and, often, blinding of those involved as to the exposure each participant is receiving, therefore reducing information bias. A unique strength of this study design is that it eliminates confounding bias, both known and unknown. Furthermore, the trial tends to be statistically efficient. If properly designed and done, a randomised controlled trial is likely to be free of bias and is thus especially useful for examination of small or moderate effects. In observational studies, bias might only account for small or moderate differences.

Randomised controlled trials have drawbacks too. External validity is one. Whereas the randomised controlled trial, if properly done, has internal validity—ie, it measures what it sets out to measure—it might not have external validity. This term indicates the extent to which results can be generalised to the broader community. Unlike the observational study, the randomised controlled trial includes only volunteers who pass through a screening process before inclusion. Those who volunteer for trials tend to be different from those who do not; for example, their health might be better. Another limitation is that a randomised controlled trial cannot be used in some instances, since intentional exposure to harmful substances—eg, toxins, bacteria, or other noxious exposures—would be unethical. As with cohort studies, the randomised controlled trial can be prohibitively expensive. Indeed, the cost of large trials runs into the tens of millions of US dollars.

Measurement of outcomes
Confusing fractions
Identification and quantification of outcomes is the business of research. However, slippery terminology often complicates matters for investigators and readers alike. For example, the term rate (as in maternal mortality rate) has been misused in textbooks and journal articles for decades. Additionally, rate is often used interchangeably with proportions and ratios. Figure 3 presents a simple approach to classification of these common terms.

Non-randomised trials: penultimate design?
Some experimental trials do not randomly allocate participants to exposures—eg, treatments or prevention strategies. Instead of using truly random techniques, investigators often use methods that fall short of the mark—eg, alternate assignment. The US Preventive Services Task Force and Canadian Task Force on the Periodic Health Examination designate this research design as class II-1, indicating less scientific rigour than randomised trials but more than analytical studies (panel).

After the investigators have assigned participants to treatment groups, the way a non-randomised trial is done and analysed resembles that of a cohort study. The exposed and unexposed are followed forward in time to ascertain the frequency of outcomes. Advantages of a non-randomised trial include use of a concurrent control group and uniform ascertainment of outcomes for both groups. However, selection bias can occur.
A ratio is a value obtained by dividing one number by another. These two numbers can be either related or unrelated. This feature—ie, relatedness of numerator and denominator—divides ratios into two groups: those in which the numerator is included in the denominator—eg, rate and proportion—and those in which it is not.

A rate measures the frequency of an event in a population. As shown in figure 3, the numerator (those with the outcome) of a rate must be contained in the denominator (those at risk of the outcome). Although all ratios feature a numerator and denominator, rates have two distinguishing characteristics: time and a multiplier. Rates indicate the time during which the outcomes occur and a multiplier, commonly to a base ten, to yield whole numbers. An example would be an incidence rate, indicating the number of new cases of disease in a population at risk over a defined interval of time—eg, 11 cases of tuberculosis per 100 000 persons per year.

Proportion is often used synonymously with rate, but the former does not have a time component. Like a rate, a proportion must have the numerator contained in the denominator. Since the numerator and denominator have the same units, these divide out, leaving a dimensionless quantity; a number without units. An example of a proportion is prevalence—eg, 27 of 100 at risk have hay fever. This number indicates how many of a population at risk have a condition at a particular time (here, 27%); since documentation of new cases over time is not involved, prevalence is more properly considered a proportion than a rate.

Although all rates and proportions are ratios, the opposite is not true. In some ratios, the numerator is not included in the denominator. Perhaps the most notorious example is the maternal mortality ratio. The definition includes women who die of pregnancy-related causes in the numerator, and women with livebirths (usually 100 000) in the denominator. However, not all those in the numerator are included in the denominator—eg, a woman who dies of an ectopic pregnancy cannot be in the denominator of women with livebirths. Thus, this venerable misnomer is actually a ratio, not a rate, a fact only recently appreciated.

Measures of association: risky business
Relative risk (also termed the risk ratio) is another useful ratio: the frequency of outcome in the exposed group divided by the frequency of outcome in the unexposed. If the frequency of the outcome is the same in both groups, then the ratio is 1·0, indicating no association between exposure and outcome. By contrast, if the outcome is more frequent in those exposed, then the ratio will be greater than 1·0, implying an increased risk associated with exposure. Conversely, if the frequency of disease is less among the exposed, then the relative risk will be less than 1·0, implying a protective effect.

Also known as the cross-products ratio or relative odds, the odds ratio has different meanings in different settings. In case-control studies, this measure is the usual measure of association. It indicates the odds of exposure among the case group divided by the odds of the exposure among controls. If cases and controls have equal odds of having the exposure, the odds ratio is 1·0, indicating no effect. If the cases have a higher odds of exposure than the controls, then the ratio is greater than 1·0, implying an increased risk associated with exposure. Similarly, odds ratios less than 1·0 indicate a protective effect.

An odds ratio can also be calculated for cross-sectional, cohort, and randomised controlled studies. Here, the disease-odds ratio is the ratio of the odds in favour of disease in the exposed versus that in the unexposed. In this context, the odds ratio has some appealing statistical features when studies are aggregated in meta-analysis, but the odds ratio does not indicate the relative risk when the proportion with the outcome is greater than 5–10%—ie, the term has little clinical relevance or meaning with higher incidence rates.

The confidence interval reflects the precision of study results. The interval provides a range of values for a variable, such as a proportion, relative risk, or odds ratio, that has a specified probability of containing the true value for the entire population from which the study sample was taken. Although 95% CIs are the most commonly used, others, such as 90%, are seen (and advocated). The wider the confidence interval, the less precision exists in the result, and vice versa. For relative risks and odds ratios, when the 95% CI does not include 1·0, the difference is significant at the usual 0·05 level. However, use of this feature of confidence intervals as a back-door means of hypothesis testing is inappropriate.

Conclusion
Understanding what kind of study has been done is a prerequisite to thoughtful reading of research. Clinical research can be divided into experimental and observational; observational studies are further categorised into those with and without a comparison group. Only studies with comparison groups allow investigators to assess possible causal associations, a fact often forgotten or ignored. Dichotomous outcomes of studies should be reported as measures of association with confidence intervals; testing null hypotheses at arbitrary p values of 0·05 has no basis in medicine and should be discouraged.

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


