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.[1,2] 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[3,4] 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.[5] Scientific illiteracy is a major failing of medical education.[6]

We have written this series of short essays on research methods for busy clinicians and active researchers. The needs of clinicians predominate; hopefully, this primer will produce more critical and thoughtful consumers of research, and thus better practitioners. The needs of clinicians overlap with those of researchers throughout the essays, but that overlap becomes most pronounced in the discussion of randomised controlled trials. For readers to assess randomised trials accurately, they should understand the relevant guidelines on the conduct of trials, emerging from methodological research. In presenting those discussions to clinicians, our essays will hopefully help researchers who do randomised trials as well.

We will cover descriptive studies, cohort studies, case-control studies, bias, and screening tests in separate articles, but will devote five articles to randomised controlled trials. This disproportion is intentional; randomised controlled trials are the gold standard in clinical research, and *The Lancet* publishes large numbers of them. Randomised controlled trials help to eliminate bias, and research has identified the important methodological elements of trials that minimise bias.[7,8] 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

Analogous to biological taxonomy, a simple hierarchy can be used to categorise most studies (panel).[9] 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

---

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.

Correspondence to: Dr David A Grimes
(e-mail: dgrimes@fhi.org)

the study type or provide sufficient detail to figure it out. A related problem is that authors sometimes incorrectly label the type of research done. Examples include calling non-randomised controlled trials randomised,[10] and labelling non-concurrent cohort studies case-control studies.[11–13] The adjective case-controlled is also sometimes (inappropriately) applied to any study with a comparison group.

Biology has animal and plant kingdoms. Similarly, clinical research has two large kingdoms: experimental and observational research. Figure 1 shows that one can quickly decide the research kingdom by noting whether the investigators assigned the exposures—eg, treatments—or whether they observed usual clinical practice.[14–18] 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.[19] An example of the latter would be a trial alternating months of liberal versus restricted access to electronic fetal monitoring for women in labour.[20]

With observational studies, which dominate the literature,[21] 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 nextdoor 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.[22] When more than one patient is described, it becomes a case-series report.[23]

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.[14,16,17] 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[24] and toxic-shock syndrome.[25] 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,[26] cross-sectional studies are done to examine the presence or absence of disease and the presence 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[27–30] continue to provide important information.
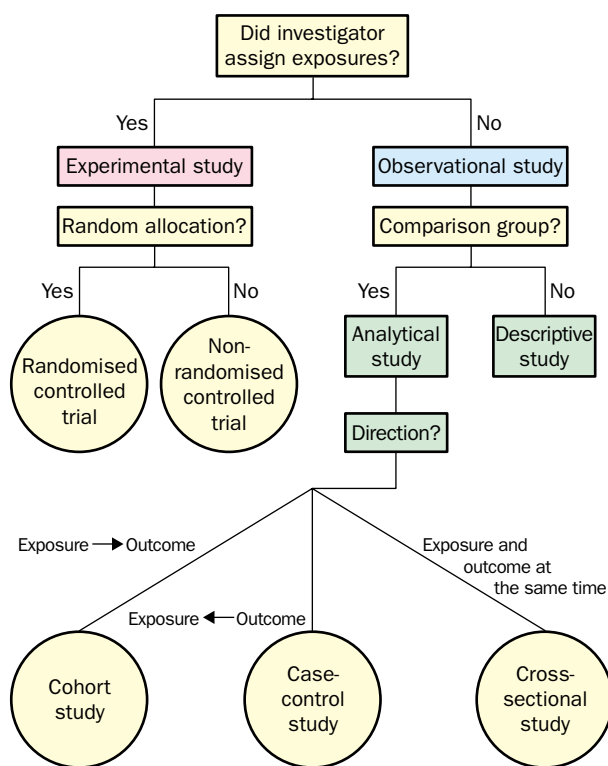

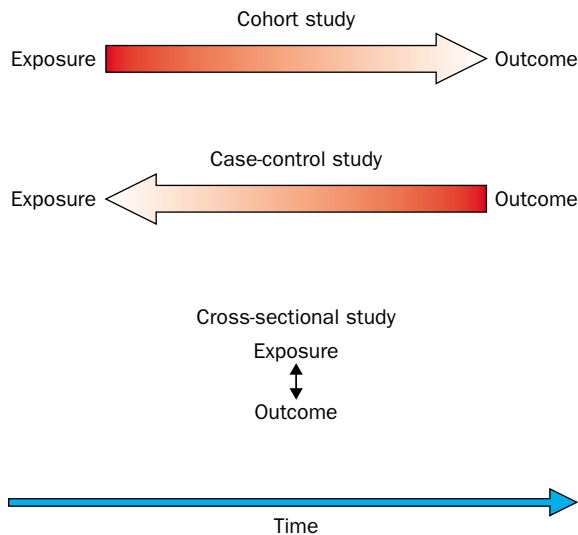Figure 1: **Algorithm for classification of types of clinical research**

Figure 2: **Schematic diagram showing temporal direction of three study designs**

Case-control study: thinking backwards

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.

Case-control studies are especially useful for outcomes that are rare or that take a long time to develop, such as cardiovascular disease and cancer. These studies often require less time, effort, and money than would cohort studies. The Achilles heel of case-control studies is choosing an appropriate control group. Controls should be similar to cases in all important respects except for not having the outcome in question. Inappropriate control groups have ruined many case-control studies and caused much harm. Additionally, recall bias (better recollection of exposures among the cases than among the controls) is a persistent difficulty in studies that rely on memory. Because the case-control study lacks denominators, investigators cannot calculate incidence rates, relative risks, or attributable risks. Instead, odds ratios are the measure of association used; when the outcome is uncommon—eg, most cancers—the odds ratio provides a good proxy for the true relative risk.

Outbreaks of food-borne diseases are a prototype for case-control studies. On a cruise ship, the entire universe of those at risk is known. Those with vomiting and diarrhoea are asked about food exposures, as are a sample of those not ill. If a higher proportion of those ill reports having eaten a food than those well, the food becomes suspect. In this way, German potato salad on a ship was linked with a serious outbreak of shigella resistant to several antibiotics.[31]

Non-randomised trial: 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.[20] The US Preventive Services Task Force[9] and Canadian Task Force on the Periodic Health Examination[32] 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.

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 easily account for small to moderate differences.[33]

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.[34] 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 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.[14] Figure 3 presents a simple approach to classification of these common terms.
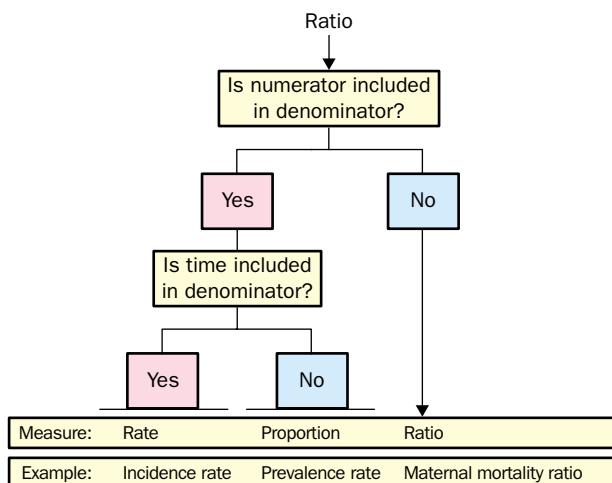
Figure 3: **Algorithm for distinguishing rates, proportions, and ratios**

A ratio is a value obtained by dividing one number by another.[26] 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.[26] 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)[26] 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,[26] 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.[35]

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).[36] 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.[36]

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.

References
1 Ramsey PG, Carline JD, Inui TS, et al. Changes over time in the knowledge base of practicing internists. *JAMA* 1991; **266:** 1103–07.
2 Evans CE, Haynes RB, Birkett NJ, et al. Does a mailed continuing education program improve physician performance? Results of a randomized trial in antihypertensive care. *JAMA* 1986; **255:** 501–04.

3   Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance: a systematic review of the effect of continuing medical education strategies. *JAMA* 1995; **274:** 700–05.

4   Sibley JC, Sackett DL, Neufeld V, Gerrard B, Rudnick KV, Fraser W. A randomized trial of continuing medical education. *N Engl J Med* 1982; **306:** 511–15.

5   Olatunbosun OA, Edouard L, Pierson RA. Physicians' attitudes toward evidence based obstetric practice: a questionnaire survey. *BMJ* 1998; **316:** 365–66.

6   Grimes DA, Bachicha JA, Learman LA. Teaching critical appraisal to medical students in obstetrics and gynecology. *Obstet Gynecol* 1998; **92:** 877–82.

7   Moher D, Schulz KF, Altman DG, Lepage L. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; **357:** 1191–94.

8   Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; **134:** 663–94.

9   US Preventive Services Task Force. Guide to clinical preventive services, 2nd edn. Baltimore, MD: Williams & Wilkins, 1996.

10 Tatum HJ, Beltran RS, Ramos R, Van Kets H, Sivin I, Schmidt FH. Immediate postplacental insertion of GYNE-T 380 and GYNE-T 380 postpartum intrauterine contraceptive devices: randomized study. *Am J Obstet Gynecol* 1996; **175:** 1231–35.

11 Silver RK, Helfand BT, Russell TL, Ragin A, Sholl JS, MacGregor SN. Multifetal reduction increases the risk of preterm delivery and fetal growth restriction in twins: a case-control study. *Fertil Steril* 1997; **67:** 30–33.

12 Garg SK, Chase HP, Marshall G, Hoops SL, Holmes DL, Jackson WE. Oral contraceptives and renal and retinal complications in young women with insulin-dependent diabetes mellitus. *JAMA* 1994; **271:** 1099–102.

13 Berenson AB, Chacko MR, Wiemann CM, Mishaw CO, Friedrich WN, Grady JJ. A case-control study of anatomic changes resulting from sexual abuse. *Am J Obstet Gynecol* 2000; **182:** 820–34.

14 Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company, 1987.

15 Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman RB, eds. Designing clinical research: an epidemiologic approach, 2nd edn. Baltimore: Lippincott Williams & Wilkins, 2001.

16 Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology, 2nd edn. New York: Oxford University Press, 1996.

17 Feinstein AR. Clinical epidemiology: the architecture of clinical research. Philadelphia: WB Saunders Company, 1985.

18 Rothman KJ. Modern epidemiology. Boston: Little, Brown and Company, 1986.

19 Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994; **272:** 125–28.

20 Leveno KJ, Cunningham FG, Nelson S, et al. A prospective comparison of selective and universal electronic fetal monitoring in 34,995 pregnancies. *N Engl J Med* 1986; **315:** 615–19.

21 Funai EF, Rosenbush EJ, Lee MJ, Del Priore G. Distribution of study designs in four major US journals of obstetrics and gynecology. *Gynecol Obstet Invest* 2001; **51:** 8–11.

22 Nerlich AG, Zink A, Szeimies U, Hagedorn HG. Ancient Egyptian prosthesis of the big toe. *Lancet* 2000; **356:** 2176–79.

23 Zink A, Rohrbach H, Szeimies U, et al. Malignant tumors in an ancient Egyptian population. *Anticancer Res* 1999; **19:** 4273–77.

24 Keys TF. A sporadic case of pneumonia due to legionnaires disease. *Mayo Clin Proc* 1977; **52:** 657–60.

25 McKenna UG, Meadows JA 3rd, Brewer NS, Wilson WR, Perrault J. Toxic shock syndrome, a newly recognized disease entity. Report of 11 cases. *Mayo Clin Proc* 1980; **55:** 663–72.

26 Last JM, ed. A dictionary of epidemiology, 2nd edn. New York: Oxford University Press, 1988.

27 Huang Z, Willett WC, Colditz GA, et al. Waist circumference, waist:hip ratio, and risk of breast cancer in the Nurses' Health Study. *Am J Epidemiol* 1999; **150:** 1316–24.

28 Kim KS, Owen WL, Williams D, Adams-Campbell LL. A comparison between BMI and Conicity index on predicting coronary heart disease: the Framingham Heart Study. *Ann Epidemiol* 2000; **10:** 424–31.

29 Hannaford PC, Kay CR. The risk of serious illness among oral contraceptive users: evidence from the RCGP's oral contraceptive study. *Br J Gen Pract* 1998; **48:** 1657–62.

30 Doll R, Peto R, Boreham J, Sutherland I. Smoking and dementia in male British doctors: prospective study. *BMJ* 2000; **320:** 1097–102.

31 Lew JF, Swerdlow DL, Dance ME, et al. An outbreak of shigellosis aboard a cruise ship caused by a multiple- antibiotic-resistant strain of Shigella flexneri. *Am J Epidemiol* 1991; **134:** 413–20.

32 Canadian Task Force on the Periodic Health Examination. The Canadian guide to clinical preventive care. Ottawa: Minister of Supply and Services Canada, 1994.

33 MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. *Lancet* 2001; **357:** 455–62.

34 Lilienfeld AM, Lilienfeld DE. Foundations of epidemiology, 2nd edn. New York: Oxford University Press, 1980.

35 Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! Evidence-based medicine 1996; **1:** 164–66.

36 Sterne JA, Smith GD. Sifting the evidence: what's wrong with significance tests? *BMJ* 2001; **322:** 226–31.