BMJ 1997;315:243-246 (26 July)

Education and debate

How to read a paper : getting your bearings (deciding what the paper is about)

Trisha Greenhalgh, *senior lecturer* [a]

[a] Unit for Evidence-Based Practice and Policy, Department of Primary Care and Population Sciences, University College London Medical School/Royal Free Hospital School of Medicine, Whittington Hospital, London N19 5NF p.greenhalgh@ucl.ac.uk

The science of "trashing" papers

It usually comes as a surprise to students to learn that some (perhaps most) published articles belong in the bin, and should certainly not be used to inform practice.[1] The first box shows some common reasons why papers are rejected by peer reviewed journals.

Why were papers rejected for publication?

- The study did not address an important scientific issue
- The study was not original (someone else had already done the same or a similar study)
- The study did not actually test the authors' hypothesis
- A different type of study should have been done
- Practical difficulties (in recruiting subjects, for example) led the authors to compromise on the original study protocol
- The sample size was too small
- The study was uncontrolled or inadequately controlled
- The statistical analysis was incorrect or inappropriate

- The authors drew unjustified conclusions from their data
- There is a significant conflict of interest (one of the authors, or a sponsor, might benefit financially from the publication of the paper and insufficient safeguards were seen to be in place to guard against bias)
- The paper is so badly written that it is incomprehensible

Most papers now appearing in medical journals are presented more or less in standard IMRAD format: Introduction (why the authors decided to do this research), Methods (how they did it, and how they analysed their results), Results (what they found), and Discussion (what the results mean). If you are deciding whether a paper is worth reading, you should do so on the design of the methods section and not on the interest of the hypothesis, the nature or potential impact of the results, or the speculation in the discussion.

Critical appraisal

The assessment of methodological quality (critical appraisal) has been covered in detail in many textbooks on evidence based medicine,[2 3 4 5 6] and in Sackett and colleagues' Users' Guides to the Medical Literature in *JAMA*.[7 8 9 10 11 12 13 14 15 16 17 18 19 20 21] If you are an experienced journal reader, the structured checklists produced by these authors will be largely self explanatory. If you are not, try these preliminary questions.

Question 1: Why was the study done, and what clinical question were the authors addressing?

The introductory sentence of a research paper should state, in a nutshell, what the background to the research is. For example, "Grommet insertion is a common procedure in children, and it has been suggested that not all operations are clinically necessary." This statement should be followed by a brief review of the published literature.

Unless it has already been covered in the introduction, the hypothesis which the authors have decided to test should be clearly stated in the methods section of the paper. If the hypothesis is presented in the negative, such as "the addition of metformin to maximal dose sulphonylurea therapy will not improve the control of type 2 diabetes," it is known as a null hypothesis.

Summary points

Many papers published in medical journals have potentially serious methodological flaws

When deciding whether a paper is valid and relevant to your practice, first establish what specific clinical question it addressed

Questions to do with drug treatment or other medical interventions should be addressed by double blind, randomised controlled trials

Questions about prognosis require longitudinal cohort studies, and those about causation require either cohort or case-control studies

Case reports, though methodologically weak, can be produced rapidly and have a place in alerting practitioners to adverse drug reactions

The authors of a study rarely actually believe their null hypothesis when they embark on their research. Being human, they have usually set out to show a difference between the two arms of their study. But the way scientists do this is to say, "Let's assume there's no difference; now let's try to disprove that theory." If you adhere to the teachings of Karl Popper, this hypotheticodeductive approach (setting up falsifiable hypotheses which you then proceed to test) is the very essence of the scientific method.[22]

Question 2: What type of study was done?

First, decide whether the paper describes a primary study, which reports research first hand, or a secondary (or integrative) one, which attempts to summarise and draw conclusions from primary studies. Primary studies, the stuff of most published research in medical journals, usually fall into one of three categories:

- Experiments, in which a manoeuvre is performed on an animal or a volunteer in artificial and controlled surroundings;
- Clinical trials, in which an intervention, such as a drug treatment, is offered to a group of patients who are then followed up to see what happens to them; or
- Surveys, in which something is measured in a group of patients, health professionals, or some other sample of individuals.

The second box shows some common jargon terms used in describing study design.

Terms used to describe design features of clinical research studies

*Parallel group comparison*Each group receives a different treatment, with both groups being entered at the same time; results are analysed by comparing groups

*Paired (or matched) comparison*Subjects receiving different treatments are matched to balance potential confounding variables such as age and sex; results are analysed in terms of differences between subject pairs

*Within subject comparison*Subjects are assessed before and after an intervention and results analysed in terms of changes within the subjects

*Single blind*Subjects did not know which treatment they were receiving

*Double blind*Neither did the investigators

*Crossover*Each subject received both the intervention and control treatments (in random order), often separated by a washout period with no treatment

*Placebo controlled*Control subjects receive a placebo (inactive pill) which should look and taste the same as the active pill. Placebo (sham) operations may also be used in trials of surgery

*Factorial design*A study which permits investigation of the effects (both separately and combined) of more than one independent variable on a given outcome (for example, a 2x2 factorial design tested the effects of placebo, aspirin alone, streptokinase alone, or aspirin plus streptokinase in acute heart attack[23])

Secondary research is made up of:

- Overviews, which may be divided into:

[Non-systematic] reviews, which summarise primary studies;

Systematic reviews, which do this according to a rigorous and predefined methodology; and

Meta-analyses, which integrate the numerical data from more than one study.

- Guidelines, which draw conclusions from primary studies about how clinicians should be behaving.
- Decision analyses, which use the results of primary studies to generate probability trees to be used by health professionals and patients in making choices about clinical management.[24] [25] [26]
- Economic analyses, which use the results of primary studies to say whether a particular course of action is a good use of resources.

Question 3: Was this design appropriate to the research?

This question is best addressed by considering what broad field of research is covered by the study. Most research studies are concerned with one or more of the broad fields shown in the box below.

Broad fields of research

- *Therapy:* testing the efficacy of drug treatments, surgical procedures, alternative methods of service delivery, or other interventions. Preferred study design is randomised controlled trial
- *Diagnosis:* demonstrating whether a new diagnostic test is valid (can we trust it?) and reliable (would we get the same results every time?). Preferred study design is cross sectional survey in which both the new test and the gold standard are performed
- *Screening:* demonstrating the value of tests which can be applied to large populations and which pick up disease at a presymptomatic stage. Preferred study design is cross sectional survey
- *Prognosis:* determining what is likely to happen to someone whose disease is picked up at an early stage. Preferred study design is longitudinal cohort study
- *Causation:* determining whether a putative harmful agent, such as environmental pollution, is related to the development of illness. Preferred study design is cohort or case-control study, depending on how rare the disease is, but case reports may also provide crucial information

Randomised controlled trials

In a randomised controlled trial, participants are randomly allocated by a process equivalent to the flip of a coin to either one intervention (such as a drug) or another (such as placebo treatment or a different drug). Both groups are followed up for a specified period and analysed in terms of outcomes defined at the outset (death, heart attack, serum cholesterol level, etc). Because, on average, the groups are identical apart from the intervention, any differences in outcome are, in theory, attributable to the intervention.

Some trials comparing an intervention group with a control group are not randomised trials. Random allocation

may be impossible, impractical, or unethical—for example, in a trial to compare the outcomes of childbirth at home and in hospital. More commonly, inexperienced investigators compare one group (such as patients on ward A) with another (such as patients on ward B). With such designs, it is far less likely that the two groups can reasonably be compared with one another on a statistical level.

A randomised controlled trial should answer questions such as the following:

- Is this drug better than placebo or a different drug for a particular disease?
- Is a leaflet better than verbal advice in helping patients make informed choices about the treatment options for a particular condition?

It should be remembered, however, that randomised trials have several disadvantages (see box).[27] Remember, too, that the results of a trial may have limited applicability as a result of exclusion criteria (rules about who may not be entered into the study), inclusion bias (selection of subjects from a group unrepresentative of everyone with the condition), refusal of certain patient groups to give consent to be included in the trial,[28] analysis of only predefined "objective" endpoints which may exclude important qualitative aspects of the intervention, and publication bias (the selective publication of positive results).[29]

Randomised controlled trial design

Advantages

- Allows rigorous evaluation of a single variable (effect of drug treatment versus placebo, for example) in a precisely defined patient group (postmenopausal women aged 50-60 years)
- Prospective design (data are collected on events that happen after you decide to do the study)
- Uses hypotheticodeductive reasoning (seeks to falsify, rather than confirm, its own hypothesis)
- Potentially eradicates bias by comparing two otherwise identical groups (but see below)
- Allows for meta-analysis (combining the numerical results of several similar trials at a later date)

Disadvantages

- Expensive and time consuming; hence, in practice:
- Many randomised controlled trials are either never done, are performed on too few patients, or are undertaken for too short a period
- Most are funded by large research bodies (university or government sponsored) or drug companies, who ultimately dictate the research agenda
- Surrogate endpoints are often used in preference to clinical outcome measures may introduce "hidden bias," especially through:
- Imperfect randomisation (see above)
- Failure to randomise all eligible patients (clinician only offers participation in the trial to patients he or she considers will respond well to the intervention)
- Failure to blind assessors to randomisation status of patients

There is now a recommended format for reporting randomised controlled trials in medical journals.[30] You should try to follow it if you are writing one up yourself.

Cohort studies

In a cohort study, two (or more) groups of people are selected on the basis of differences in their exposure to a particular agent (such as a vaccine, a drug, or an environmental toxin), and followed up to see how many in each group develop a particular disease or other outcome. The follow up period in cohort studies is generally measured in years (and sometimes in decades), since that is how long many diseases, especially cancer, take to develop. Note that randomised controlled trials are usually begun on patients (people who already have a disease), whereas most cohort studies are begun on subjects who may or may not develop disease.

PETER BROWN

View larger version (145K):
[in this window]
[in a new window]

A special type of cohort study may also be used to determine the prognosis of a disease (what is likely to happen to someone who has it). A group of patients who have all been diagnosed as having an early stage of the disease or a positive result on a screening test is assembled (the inception cohort) and followed up on repeated occasions to see the incidence (new cases per year) and time course of different outcomes.

The world's most famous cohort study, which won its two original authors a knighthood, was undertaken by Sir Austin Bradford Hill, Sir Richard Doll, and, latterly, Richard Peto. They followed up 40 000 British doctors divided into four cohorts (non-smokers, and light, moderate, and heavy smokers) using both all cause mortality (any death) and cause specific mortality (death from a particular disease) as outcome measures. Publication of their 10 year interim results in 1964, which showed a substantial excess in both lung cancer mortality and all cause mortality in smokers, with a "dose-response" relation (the more you smoke, the worse your chances of getting lung cancer), went a long way to showing that the link between smoking and ill health was causal rather than coincidental.[31] The 20 year and 40 year results of this momentous study (which achieved an impressive 94% follow up of those recruited in 1951 and not known to have died) illustrate both the perils of smoking and the strength of evidence that can be obtained from a properly conducted cohort study.[32] [33]

A cohort study should be used to address clinical questions such as:

- Does high blood pressure get better over time?
- What happens to infants who have been born very prematurely, in terms of subsequent physical development and educational achievement?

Case-control studies

In a case-control study, patients with a particular disease or condition are identified and "matched" with controls (patients with some other disease, the general population, neighbours, or relatives). Data are then collected (for example, by searching back through these people's medical records or by asking them to recall their own history) on past exposure to a possible causal agent for the disease. Like cohort studies, case-control studies are generally concerned with the aetiology of a disease (what causes it) rather than its treatment. They lie lower down the hierarchy of evidence (see below), but this design is usually the only option for studying rare conditions. An important source of difficulty (and potential bias) in a case-control study is the precise definition of who counts as a "case," since one misallocated subject may substantially influence the results. In addition, such a design cannot show causality—the association of A with B in a case-control study does not prove that A has caused B.

A case-control study should be used to address clinical questions such as:

- Does the prone sleeping position increase the risk of cot death (the sudden infant death syndrome)?
- Does whooping cough vaccine cause brain damage?
- Do overhead power cables cause leukaemia?

Cross sectional surveys

We have probably all been asked to take part in a survey, even if only one asking us which brand of toothpaste we prefer. Surveys conducted by epidemiologists are run along the same lines: a representative sample of subjects (or patients) is interviewed, examined, or otherwise studied to gain answers to a specific clinical question. In cross sectional surveys, data are collected at a single time but may refer retrospectively to experiences in the past—such as the study of casenotes to see how often patients' blood pressure has been recorded in the past five years.

A cross sectional survey should be used to address clinical questions such as:

- What is the "normal" height of a 3 year old child?
- What do psychiatric nurses believe about the value of electroconvulsive therapy in severe depression?
- Is it true that half of all cases of diabetes are undiagnosed?

A memorable example of a case report

A doctor notices that two newborn babies in his hospital have absent limbs (phocomelia). Both mothers had taken a new drug (thalidomide) in early pregnancy. The doctor wishes to alert his colleagues worldwide to the possibility of drug related damage as quickly as possible.[35]

Case reports

Top
The science of "trashing"...
Critical appraisal
Randomised controlled trials
Cohort studies
Case-control studies
Cross sectional surveys
Case reports
The hierarchy of evidence
References

A case report describes the medical history of a single patient in the form of a story: "Mrs B is a 54 year old secretary who developed chest pain in June 1995...." Case reports are often run together to form a case series, in which the medical histories of more than one patient with a particular condition are described to illustrate an aspect of the condition, the treatment, or, most commonly these days, adverse reaction to treatment. Although this type of research is traditionally considered to be "quick and dirty" evidence, a great deal of information can be conveyed in a case report that would be lost in a clinical trial or survey .[34]

The hierarchy of evidence

Top
The science of "trashing"...
Critical appraisal
Randomised controlled trials
Cohort studies
Case-control studies
Cross sectional surveys
Case reports
The hierarchy of evidence
References

Standard notation for the relative weight carried by the different types of primary study when making decisions about clinical interventions (the "hierarchy of evidence") puts them in the following order[36]:

1. Systematic reviews and meta-analyses
2. Randomised controlled trials with definitive results (confidence intervals that do not overlap the threshold clinically significant effect)
3. Randomised controlled trials with non-definitive results (a point estimate that suggests a clinically significant effect but with confidence intervals overlapping the threshold for this effect)
4. Cohort studies
5. Case-control studies
6. Cross sectional surveys
7. Case reports.

The articles in this series are excerpts from *How to read a paper: the basics of evidence based medicine*. The book includes chapters on searching the literature and implementing evidence based findings. It can be ordered from the BMJ Bookshop: tel 0171 383 6185/6245; fax 0171 383 6662. Price £13.95 UK members, £14.95 non-members.

References

Top
The science of "trashing"...

1. Altman DG. The scandal of poor medical research. *BMJ* 1994;308:283-4.

- Critical appraisal
- Randomised controlled trials
- Cohort studies
- Case-control studies
- Cross sectional surveys
- Case reports
- The hierarchy of evidence
- References

[Free Full Text]

2. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. London, Little, Brown, 1991.
3. Sackett DL, Richardson WS, Rosenberg WMC, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. London: Churchill-Livingstone, 1996.
4. Crombie IM. *The pocket guide to critical appraisal*. London: BMJ Publishing Group, 1996.
5. Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology: the essentials*. 3rd ed. Baltimore: Williams and Williams, 1996.
6. Rose G, Barker DJP. *Epidemiology for the uninitiated*. 3rd ed. London: BMJ Books, 1993.
7. Oxman AD, Sackett DS, Guyatt GH. Users' guides to the medical literature. I. How to get started. *JAMA* 1993;270:2093-5.
8. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA* 1993;270:2598-601.
9. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *JAMA* 1994;271:59-63.
10. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? *JAMA* 1994;271:389-91. [Medline]
11. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What were the results and will they help me in caring for my patients? *JAMA* 1994;271:703-7.
12. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guides to the medical literature. IV. How to use an article about harm. *JAMA* 1994;271:1615-9. [Medline]
13. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. *JAMA* 1994;271:234-7. [Abstract]
14. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. *JAMA* 1994;272:1367-71.
15. Richardson WS, Detsky AS. Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid? *JAMA* 1995;273:1292-5. [Medline]
16. Richardson WS, Detsky AS. Users' guides to the medical literature. VII. How to use a clinical decision analysis. B. What are the results and will they help me in caring for my patients? *JAMA* 1995;273:1610-3.
17. Hayward RSA, Wilson MC, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? *JAMA* 1995;274:570-4.
18. Wilson MC, Hayward RS, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. B. Will the recommendations help me in caring for my patients? *JAMA* 1995;274:1630-2. [Medline]
19. Naylor CD, Guyatt GH. Users' guides to the medical literature. XI. How to use an article about a clinical utilization review. *JAMA* 1996;275:1435-9. [Medline]
20. Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? *JAMA* 1997;277:1552-7. [Medline]
21. O'Brien BJ, Heyland D, Richardson WS, Levine M, Drummond MF. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? *JAMA* 1997;277:1802-6. [Medline]
22. Popper K. *Conjectures and refutations: the growth of scientific knowledge*. New York: Routledge and Kegan Paul, 1963.
23. Randomised trial of intravenous streptokinase, aspirin, both, or neither among 17187 cases of suspected acute myocardial infarction: ISIS-2. (ISIS-2 Collaborative Group). *Lancet* 1988;ii:349-60.
24. Thornton JG, Lilford RJ, Johnson N. Decision analysis in medicine. *BMJ* 1992;304:1099-103. [Medline]
25. Thornton JG, Lilford RJ. Decision analysis for medical managers. *BMJ* 1995;310:791-4. [Free Full Text]
26. Dowie J. "Evidence-based", "cost-effective", and "preference-driven" medicine. *J Health Serv Res Policy* 1996;1:104-13.
27. Bero LA, Rennie D. Influences on the quality of published drug studies. *Int J Health Technology Assessment* 1996;12:209-37.

28. MacIntyre IMC. Tribulations for clinical trials. Poor recruitment is hampering research. *BMJ* 1991;302:1099-100.
29. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72.
30. Altman D. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ* 1996; 313:570-1.
31. Doll R, Hill AB. Mortality in relation to smoking: ten years' observations on British doctors. *BMJ* 1964;i:1399-414, 1460-7.
32. Doll R, Peto R. Mortality in relation to smoking: ten years' observations on British doctors. *BMJ* 1976;ii:1525-36.
33. Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ* 1994;309:901-11. [Abstract/Free Full Text]
34. MacNaughton J. Anecdotes and empiricism. *Br J Gen Pract* 1995; 45:571-2.
35. McBride WG. Thalidomide and congenital abnormalities. *Lancet* 1961;ii:1358.
36. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. *JAMA* 1995;274:1800-4.

This article has been cited by other articles:

T. **Greenhalgh** and R. Taylor

How to read a paper: Papers that go beyond numbers (qualitative research)

BMJ, September 20, 1997; 315(7110): 740 - 743.

[Full Text]