How to read a paper

Getting your bearings (deciding what the paper is about)

Trisha Greenhalgh

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.

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,²³ and in Sackett and colleagues’ Users’ Guides to the Medical Literature in JAMA.²⁴ 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.

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 hypothetico-deductive approach (setting up falsifiable hypotheses which you then proceed to test) is the very essence of the scientific method.²²

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

This is the second of 10 articles introducing non-experts to finding medical articles and assessing their value

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

Trisha Greenhalgh, senior lecturer
p.greenhalgh@ucl.ac.uk

BMJ 1997;315:243-4

Reference

¹ Greenhalgh T. BMJ 1997;315:243-4


³ Greenhalgh T. BMJ 1997;315:243-4

⁴ Greenhalgh T. BMJ 1997;315:243-4

⁵ Greenhalgh T. BMJ 1997;315:243-4

⁶ Greenhalgh T. BMJ 1997;315:243-4

⁷ Greenhalgh T. BMJ 1997;315:243-4

⁸ Greenhalgh T. BMJ 1997;315:243-4

⁹ Greenhalgh T. BMJ 1997;315:243-4

¹⁰ Greenhalgh T. BMJ 1997;315:243-4

¹¹ Greenhalgh T. BMJ 1997;315:243-4

¹² Greenhalgh T. BMJ 1997;315:243-4

¹³ Greenhalgh T. BMJ 1997;315:243-4

¹⁴ Greenhalgh T. BMJ 1997;315:243-4

¹⁵ Greenhalgh T. BMJ 1997;315:243-4

¹⁶ Greenhalgh T. BMJ 1997;315:243-4

¹⁷ Greenhalgh T. BMJ 1997;315:243-4

¹⁸ Greenhalgh T. BMJ 1997;315:243-4

¹⁹ Greenhalgh T. BMJ 1997;315:243-4

²⁰ Greenhalgh T. BMJ 1997;315:243-4

²¹ Greenhalgh T. BMJ 1997;315:243-4

²² Greenhalgh T. BMJ 1997;315:243-4

²³ Greenhalgh T. BMJ 1997;315:243-4

²⁴ Greenhalgh T. BMJ 1997;315:243-4

²⁵ Greenhalgh T. BMJ 1997;315:243-4

²⁶ Greenhalgh T. BMJ 1997;315:243-4

²⁷ Greenhalgh T. BMJ 1997;315:243-4

²⁸ Greenhalgh T. BMJ 1997;315:243-4

²⁹ Greenhalgh T. BMJ 1997;315:243-4

³⁰ Greenhalgh T. BMJ 1997;315:243-4

³¹ Greenhalgh T. BMJ 1997;315:243-4

³² Greenhalgh T. BMJ 1997;315:243-4

³³ Greenhalgh T. BMJ 1997;315:243-4

³⁴ Greenhalgh T. BMJ 1997;315:243-4

³⁵ Greenhalgh T. BMJ 1997;315:243-4

³⁶ Greenhalgh T. BMJ 1997;315:243-4

³⁷ Greenhalgh T. BMJ 1997;315:243-4

³⁸ Greenhalgh T. BMJ 1997;315:243-4

³⁹ Greenhalgh T. BMJ 1997;315:243-4

⁰ Greenhalgh T. BMJ 1997;315:243-4

⁹ Greenhalgh T. BMJ 1997;315:243-4
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. Moreover, 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). 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, analysis of only predefined “objective” endpoints which may exclude important qualitative aspects of the intervention, and publication bias (the selective publication of positive results).

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 provide crucial information
There is now a recommended format for reporting randomised controlled trials in medical journals. 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.

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.3233

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 linking 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...
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.1

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?

Case reports

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 1955…” 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.2

The hierarchy of evidence

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:3

(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.

References

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? BMJ 1993;270:259-61.
9. Guyatt GH, Sackett DL, Cook DJ: Users’ guides to the medical literature. III. 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;271:1116-8.
16. Richardson WS, Denke AS: Users’ guides to the medical literature. VII. How to use a clinical decision analysis. B. What were the results and will they help me in caring for my patients? JAMA 1995;273:1610-5.
21. O’Brien BJ, Hayland 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 were the results and will they help me in caring for my patients? JAMA 1997;277:1802-6.