

EVIDENCE-BASED MEDICINE

(1) RESEARCH METHODS

By Claire Jones, DipPresSci, MRPharmS

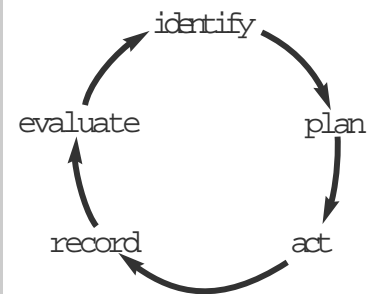
Pharmacists are constantly bombarded with the latest evidence for the efficacy of new drugs, the diagnoses of diseases or their causes. Indeed, we have a professional obligation to keep up to date. This first article of two is a basic guide to research methods and revises some of the common terms pharmacists may come across when reading medical literature



identify gaps in your knowledge

1. Read the news item about type 1 diabetes (p830). Do you understand the meaning of the term "relative risk 0.96"?
2. Can you list three types of research method?
3. What type of research method carries the most weight?

This article relates to the Royal Pharmaceutical Society's core competencies of "information systems and sources of information" and "decision making in relation to drug therapy" (see "Medicines, ethics and practice — a guide for pharmacists", number 25, July 2001, p104). You should consider how it will be of value to your practice.



Evidence-based medicine is a term that pharmacists are encountering with increasing frequency, particularly with the new roles being developed (eg, formulary pharmacists and positions on primary care trust prescribing sub-committees). Basically, the term means using good evidence to make sound clinical decisions. Valid evidence of clinical benefit and cost-effectiveness is what will influence purchasers of health care. So we come to two fundamental questions: "what is the best evidence?" and "how does one know if a piece of research is good enough to be relied on?"

"THE HIERARCHY OF EVIDENCE"

The term "hierarchy of evidence" is one you have probably come across before. It simply describes how different types of evidence are ranked in terms of importance when decisions need to be made about clinical interventions. Evidence can be from a primary or secondary source. Primary evidence is direct research (eg, a randomised control trial) whereas secondary evidence analyses primary evidence (eg, meta-analysis). Figure 1 (p840) shows the evidence pyramid. The following describes each type of trial or report and uses heart failure to illustrate its place in research.

Systematic reviews or meta-analyses Systematic reviews and meta-analyses are the types of evidence at the pinnacle of the pyramid. A medical literature review usually involves an expert in a particular field collating and assessing existing knowledge in that field and making a summary. However, this type of review often lacks rigour and is vulnerable to the reviewer's personal opinions. Other inadequacies include the omission of small but significant studies.

Systematic review is a formal method of review that requires certain steps to be followed in order to overcome such problems. For example, all sources of literature with respect to a specific question on heart failure would be carefully searched for all trials (both published and unpublished), each trial found would be assessed by independent reviewers and the results combined and discussed before inferences could be made. A "Cochrane systematic review" is a type of review aimed at giving evidence in a health care setting.

A meta-analysis is a specific statistical strategy for deriving a single numerical estimate from the results of several randomised con-

trolled trials (RCTs). This means that the problem of small sample sizes in some trials (eg, results that are unrepresentative of a population or results that can be easily obscured by a rare occurrence) can be overcome because the pooling of trials increases overall sample size. For example, the results of all RCTs on the effects of an angiotensin-converting enzyme inhibitor could be pooled.

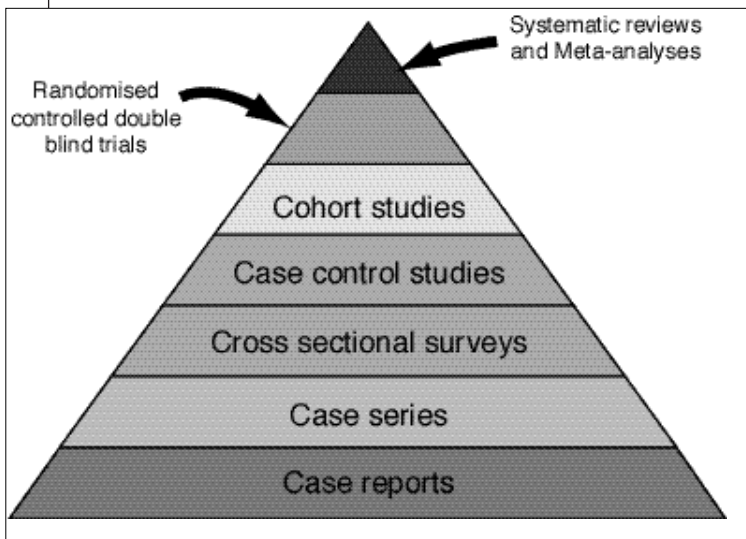
It must be stressed, however, that although these two types of evidence are thought to carry the most weight, they are not immune to flaws and every review or analysis must also be critically appraised before it can be used with confidence.

Randomised controlled trials An RCT is the type of research normally used to assess the relative effect of drugs. For example, an RCT might be designed to investigate if drug X is better than placebo or drug Y for the treatment of heart failure. Typically, patients are randomly allocated to one intervention or another (ie, split into two groups). One group is a control group. The two patient groups should be identical except for the treatment they have been randomised to receive. Only look for other types of trial if you cannot find an RCT.

Cohort studies Cohort studies are observational studies of subjects with a specific disease or characteristic (eg, treatment with drug X) who are followed over a period (usually years) to detect complications or new events. The group may be compared with a control group. Studies are generally concerned with what causes a disease. Problems include the time a study can take and the influence of other lifestyle variables. Going back to our heart failure example, all patients newly diagnosed with heart failure registered in a number of general practices could be observed to determine percentage survival over time.

Case control studies Case control studies are also concerned with what causes a disease. Patients with a particular disease are matched with controls (people without the disease), but rather than following

Ms Jones is the assistant head of NHS service development at the National Pharmaceutical Association and a National Prescribing Centre training adviser



the subjects into the future, data on past exposure to possible causal agents are collected (eg, by searching through medical records or interviewing subjects). Histories are then compared. So patients with heart failure would be matched with patients without in order to determine what risk factors might be linked to the development of heart failure. This type of research will produce results faster than cohort studies, but it is not as reliable. Even if a disease and a condition are related statistically, this does not mean that the condition caused the disease.

Cross sectional surveys A cross sectional survey is a measure of the frequency of a disease or risk factor in a defined population at a given time. For example, the records of all patients in a number of general practices could be reviewed to determine the number of patients with heart failure (and therefore the prevalence of heart failure in that population) or the percentage of patients with heart failure being prescribed an ACE inhibitor.

Case reports and case series A case report describes the medical history of a single patient in the form of a story and a case series is a collection of similar reports. They are usually used to record and/or alert other health professionals to rare occurrences. For example, if a patient who has taken two different drugs separately in the past takes them together and develops a life threatening arrhythmia and the doctor treating the patient suspects that the two drugs may be interacting, he or she could produce a case report. Because there is no control group, case reports and series are not valid statistically. They provide anecdotal evidence.

COMMON TERMS

Once you have a basic idea of what each type of study involves and its position in the hierarchy of evidence, you may find that you need to look at a study in more detail to judge whether it can be used for your purposes (ie, critical appraisal). To do this, you need an understanding of the terms and phrases used to describe the design features of a study. This section revises some common terms.¹⁻⁴

Baseline characteristic/value The terms baseline characteristic and baseline value are used for a variable that is measured or assessed before an intervention is started (“baseline examination”). A set of collected variables is called baseline data.

Controls If a trial is controlled it simply means that the treatment group is compared with another group. This looks at “between subject variation” ie, differences between subjects. There are two basic types of control:

- 1 Placebo: control patients receive an inactive “treatment” which should look, feel and taste the same as the active intervention given to the treatment group (ie, a “matching placebo”)

- 1 Active: the control is a treatment that is already in use because the use of a placebo (non treatment) may be unethical

Trial groups In most trials, subjects will be split into different groups, also known as “arms”. RCTs usually have two arms (eg, control and treatment), but some may have more. Other terms used to describe groupings are:

- 1 Parallel groups: subjects are randomised to one of two different interventions (eg, medicine and placebo), usually for the entire trial and then the two groups are compared ie, a typical RCT
- 1 Matching/pairing: investigators identify pairs of subjects who have identical characteristics (eg, weight, sex, age) before randomising them so that each one receives a different intervention. This will reduce the influence of factors such as age on the results
- 1 Cross over: each patient receives two sequential interventions (eg, treatment and control) in random order, separated by a “washout period” (a period of no treatment). This design feature will allow “within subject differences” to be investigated. Because the same subject is used, there is no need to for matching
- 1 Factorial: a study with a factorial design allows the effects of treatments used in combination as well as separately to be investigated. A “2x2” design has four groups, for example, placebo group, drug X, drug Y and drug X plus drug Y

Randomisation In an RCT patients are randomised to either the intervention or the control group. The method of randomisation is important. The most robust kinds of randomisation use computer-generated numbers or random number tables. Randomisation by the toss of a coin or by sequential allocation is not acceptable.

Blinding Blinding is a term used to describe conditions that are imposed to keep groups of individuals from knowing which subjects have or have not received an intervention. Sometimes called masking, blinding is used to reduce bias and the placebo effect. Different types of blinding may be possible:

- 1 Single blind: the patients do not know which intervention they receive
- 1 Double blind: neither the patients nor the investigators know who was receiving which treatment
- 1 Triple blind: neither the patients nor the investigators nor the data analysers know who was receiving which treatment
- 1 Open (label) study: no blinding. But note that open label trials can still be randomised.

Endpoint, event, primary outcome The variable used to judge the effectiveness of an intervention (eg, response rate) can be described by various terms: endpoint, event, primary outcome. It is essential that this variable is defined in advance. A “hard endpoint” is an explicit outcome variable that is not vulnerable to serious errors in measurement or interpretation (eg, death).

STATISTICS

An explanation of the various statistical tests used in research are outside the scope of this article, but the two main terms that should be understood are p-value and confidence intervals. How the results are analysed is also significant.

P-value The P-value is the result of the statistical test used to assess the probability that the result of the trial is a real effect and did not occur purely by chance. By convention a P-value below 0.05 (a one in 20 likelihood that the result occurred by chance) is accepted as indicating a true difference and is described as a statistically significant result.

95 per cent confidence interval (CI) Trials have some degree of uncertainty because a result from a trial on a sample would not be exactly the same if the intervention were applied to a whole population. The CI around the result represents the range of values within which the true population value lies. By convention 95 per cent CI

are used, which means that you can be 95 per cent sure that the true result lies between a certain range.

Intention to treat analysis An intention to treat analysis means that the results used include all the original patients, including those who have dropped out of the trial. This analysis more closely reflects a real life situation where some patients are not compliant with therapy.

HOW THE RESULTS ARE EXPRESSED

The benefits or harms of a treatment can be shown in various ways:

- 1 Drug X produced an absolute reduction in deaths by 7 per cent (“absolute risk reduction”)
- 1 Drug X reduced the death rate by 28 per cent (“relative risk reduction”)
- 1 Drug X increased the patients’ survival rate from 75 to 82 per cent
- 1 14 people would need to be treated with drug X to prevent one death (“number needed to treat”)

Which result sounds the most impressive? In fact, all of the above relate to the same results (listed in Panel 1 below). This illustrates that the way in which results are presented may affect how they are perceived. An understanding of the principles underlying the expression of results in terms of relative or absolute risks is invaluable when assessing studies. In practice, the “number needed to treat” is the most useful expression of results. This section attempts to explain the four common terms used to express results and to show how they are calculated, using the figures in Panel 1.

Absolute risk reduction The absolute risk reduction (ARR) is the absolute amount by which drug X reduces the risk of death, calculated as:

$$\begin{aligned} \text{ARR} &= (\text{event rate in control group} - \text{event rate in intervention group}) \times 100 \\ &= (501/1,998 - 358/2,000) \times 100 \\ &= 7 \text{ per cent} \end{aligned}$$

ie, drug X produced an absolute reduction in death by 7 per cent

Relative risk reduction The relative risk of an outcome is the chances of that outcome occurring in the treatment group compared with the chances of it occurring in the control group. If the chances are the same in both groups, the relative risk is 1. The relative risk reduction (RRR) is the amount by which the risk (of death) is reduced by drug X as a comparative percentage of the control, calculated as:

$$\begin{aligned} \text{RRR} &= [(\text{event rate in control group} - \text{event rate in intervention group}) \times 100] \div \text{event rate in control group} \\ &= [(501/1,998 - 358/2,000) \times 100] \div 501/1,998 \\ &= 28 \text{ per cent} \end{aligned}$$

ie, drug X reduced the death rate by 28 per cent

Note that relative comparisons will make the results sound more impressive and this is a tactic often used by manufacturers. Similarly, absolute comparisons may be used to make the risk of side effects sound smaller. Be aware that both absolute and relative risks may be used together in the same set of results, so make sure you are clear on what type of expression is being used.

PANEL 1: FIGURES FOR EXAMPLE CALCULATIONS

Groups	Total number of patients randomised to each group	Outcome at five years
Intervention group (received drug X)	2,000	358 dead 1,642 alive
Control group (received placebo)	1,998	501 dead 1,497 alive

action : practice points

- 1. Explain the term “number needed to treat” to a colleague.
- 2. Do you really understand the difference between relative and absolute risk? If not, go to www.breastcancer.org/treatment_stats.html and read “Weighing treatment pros and cons”.
- 3. Explore the sources of information available to help you with evidence-based medicine. A starting point is to visit www.evidence-based-medicine.co.uk and look at p4 of the document titled “What is evidence-based medicine?”

evaluate

How could your learning have been more effective?
What will you do now and how will this be achieved?

Number needed to treat The “number needed to treat” (NNT) is the number of people who need to be treated to produce one additional successful outcome. For example, how many patients need to be treated by drug X instead of placebo to prevent one death at five years?

$$\begin{aligned} \text{NNT} &= 100/\text{ARR} \\ &= 100/7 \\ &= 14 \end{aligned}$$

ie, 14 people would need to be treated with drug X to prevent one death at five years

Odds ratio The odds of an event compares the probability of the event occurring with the probability that it will not occur. If the odds are greater (or less) than 1, an event is more (or less) likely to happen. The odds ratio (OR) is the ratio of patients in the treatment group subscribing to a particular end point of the trial to the number who do not, compared with the equivalent patients in the control group.

The odds of dying compared with the odds of surviving for patients in the intervention group treated with drug X are 358/1,642 = 0.22, and for patients in the control group 501/1,497 = 0.33. The OR will therefore be 0.22/0.33 = 0.67.

An OR of 1 would mean that drug X had no effect ie, there was no overall difference in outcomes between the intervention and control group. Sometimes, the terms “percentage reduction in odds ratio” is used.

CONCLUSION

I am sure that for the majority of pharmacists, this article is revision. For those of you who are less experienced with reading and appraising trial evidence I hope this has given you a taster of how to begin to understand the different types of trial evidence and the common terms and phrases that are used in them. The next article in this series will describe how to go about critically appraising a randomised controlled trial and where to start in getting some formal training to keep abreast of the good independent sources of evidence. In the ever-changing world of evidence-based medicine it is essential that pharmacists have the knowledge and skills to play a part in reviewing and appraising different types of trial evidence.

REFERENCES

- 1. Greenhalgh. How to read a paper: the basics of evidence based medicine. *BMJ* 1997;315:480-1.
- 2. NHS Critical Appraisal Skills Programme (CASP) available at www.phru.org.uk/~casp/index.htm (accessed 11 June 2002)
- 3. Evidence based medicine. *MeReC Bull* 1995;6: 45-8.
- 4. An introduction to assessing medical literature. *MeReC Briefing* 1995;9:1-8.