**WeMeReC**
Welsh Medicines Resource Centre
Canolfan Adnoddau Moddion Cymru

# Evaluating clinical evidence - what do I need to know?

Commentary on the latest evidence for medical interventions is reported on, not only in the medical press, but increasingly in the popular press as well. Results of published medical studies routinely appear in newspapers and magazines under sensational headlines. The availability of information via the internet means that such commentary is not only widely accessible, but is often read by patients before healthcare professionals within the NHS have time to familiarise themselves with the research and act accordingly.

It is impractical, indeed unnecessary, for one individual to critically assess all available evidence. To this end, there are many excellent sources of secondary information that evaluate research findings, including those studies "hitting the headlines". Many of these are listed in the *Resources* section on page 5.

Nevertheless, a basic knowledge of how studies are evaluated is crucial to understanding summaries that are compiled – to judging whether a study is relevant to your practice and to knowing how study results might apply to your patients. Such knowledge is essential if you want to discuss the benefits and risks of therapies with patients.



Experience in teaching evaluation skills to healthcare professionals has revealed that there are some key aspects of clinical evidence that are not well understood. This bulletin may be helpful to you if you are unable to answer "yes" to any of the following questions:

♦ Are you confident explaining the difference between absolute and relative risk reductions?

♦ Can you calculate a number-needed-to-treat (NNT)?

♦ Do you know what the advantages and disadvantages of using surrogate or composite endpoints are?

♦ Do you understand what the power of a study refers to?

This bulletin aims to cover some of the basic principles for successful evaluation of clinical intervention studies. The associated case-based education module provides an opportunity to practise applying these.

Factors to consider when evaluating studies fall under four broad headings:

1. Is the evidence relevant – how does it apply to my practice?
2. Is the study valid – is the research sound?
3. What are the results and what do they mean – do I understand the numbers?
4. How will this affect my practice? What are the implications for my patients?

An understanding of these issues will be reflected in your interactions with patients and will allow well informed discussions about the pros and cons of initiating or continuing therapies.

NHS WALES
GIG CYMRU

## Is the evidence relevant?

*Is it helpful – does it matter?*

Not everything that is categorised as "news" is information that you need to know. Before reading a paper or a summary of evidence, you will make an assessment as to whether it will be a good use of your time. The usefulness of the exercise will depend on whether the evidence is applicable to situations you encounter in clinical practice and, if so, how often.[1] It will also depend on whether the information relates to events that matter to your patients.[1,2] When assessing studies it is necessary to ask, "what, where, in whom, and why"?

*Who were the study participants?*

Ask yourself whether the subjects who participated in the trial are reasonably similar to the patients who would be offered the treatment in your clinical practice. How were they recruited? What were the exclusion criteria? Factors such as age, gender, diagnosis, and co-morbidity are important.

*Where was the study conducted and for how long?*

Where a study is conducted will have a bearing on, not only the population/s from which subjects were drawn, but the environmental factors they will have been exposed to, and the healthcare available to them. Consider whether participants were observed in a primary care setting or were hospitalised. The length of a study should be noted, alongside other factors such as whether subjects were followed-up and, if so, how frequently. Was the study long enough to allow relevant clinical events to be detected? Check when the study began. Has practice changed in the intervening period?

*Were the interventions appropriate?*

Was the level of intervention or the dose of medication adequate? If the study was controlled, was the comparator suitable? Is similar treatment for your patients feasible? Can you deliver the same interventions or are they accessible elsewhere?

---

**Things to look for in a study...**

♦ Exclusion criteria
♦ Time-scales
♦ Appropriate comparators
♦ Randomisation, concealed allocation, and double-blind design
♦ Power calculations
♦ Valid, relevant endpoints
   Check for **patient-oriented** rather than **disease-oriented outcomes** (e.g. pain relief vs reductions in inflammatory markers).

---

## Is the study valid?

*Was the study design appropriate?*

The choice of study design will depend on the research being undertaken. Study designs can be roughly divided into two groups:

♦ Experimental or interventional e.g. randomised controlled trials – these are considered the "gold standard" for assessing the effectiveness of interventions.

♦ Observational or non-interventional e.g. cohort, case-control, and cross-sectional studies. Broadly speaking, these designs are more appropriate for studies of diagnosis, screening, prognosis, or causation.

Each study design has strengths and weaknesses. In this bulletin we focus on appraising randomised controlled trials. Conducting randomised trials can minimise accidental or intentional bias, but does not automatically guarantee quality. The methods used in a study will determine how useful the results are.[3]

*Was the allocation of subjects into control and intervention groups truly random?*

Participants should be randomly assigned to a group by a process equivalent to the flip of a coin. Each group is allotted a different intervention, one of which will act as a control. If truly random, this method of allocation should help to ensure that, on average, the groups are identical apart from the interventions.[4] Any differences in outcome can then, in theory, be attributed to the interventions.

*Was allocation concealed from investigators?*

It is important to consider whether the people enrolling subjects in a trial had any way of knowing which group a potential participant would be assigned to. Inadequate concealment of treatment allocation can have a profound effect on recruitment to a study. Studies without concealed allocation consistently overestimate benefit.[5]

*Did the study have enough participants to minimise the play of chance?*

This is often referred to as the **power** of a study. The power of a study is the probability that the trial will detect a significant difference in treatment outcomes if a real difference exists. By arbitrary convention, a power of 0.8 is generally accepted as being adequate in most research studies. Prior to commencing a study, a calculation should be done to determine how large the **sample size**, or study population, needs to be in order to achieve this degree of power.[6]

*Was it a double-blind study?*

Ideally, participants and outcome assessors should be blind to the nature of the assigned treatment. Blinding reduces bias in the way clinicians observe the outcomes of treatment (observer bias) and the way in which participants report their symptoms.[7]

*Is the trial endpoint clearly defined? Are multiple outcomes measured? Are surrogate or composite endpoints used?*

Investigators should focus on the smallest number of clinical endpoints feasible. If they specify more than one endpoint, they should do so at the outset of the study. The study should be adequately powered for the primary endpoints and results for all endpoints should be reported.

A **surrogate endpoint** is an outcome, which is relatively easy to measure, such as a biological marker, laboratory finding, or physical sign, that is often used when observing important clinical events is expected to be impractical or expensive (frequently involving too long a follow-up period). Examples of surrogate outcomes include bone mineral density, endoscopy findings, and cholesterol concentrations. The validity of a surrogate endpoint depends on the extent to which it correlates with, or is indicative or predictive of, the relevant clinical outcome (e.g. vertebral fracture, gastrointestinal bleed, myocardial infarction, or death).[4]

A **composite endpoint** is an outcome that has more than one component. For example, a composite cardiovascular endpoint could be defined as myocardial infarction, or stroke, or death occurring in a subject. Using composite endpoints increases event rates, thus reducing the number of subjects required for studies to be adequately powered. However, the validity of a composite endpoint is dependent on the components being similar in the way they are affected by a treatment, the frequency with which they occur, and their importance to patients.[8] Consider the composite endpoint of hospitalisation or death from pneumonia. Do both components of this endpoint warrant equal weighting? Be watchful for reporting of components as discrete endpoints.

*Were all participants treated in the same way?*

Participants, regardless of the interventions they are assigned to receive, should be treated identically in every other respect so that any differences in outcomes between groups can be ascribed to the interventions with as much confidence as possible. Consider whether adjunctive therapy (such as counselling or diets) or other support (such telephone interviews or follow-up visits) was provided.

## What are the results?

*What are the facts?*

In processing any report, whatever its source, it is important to look past the presentation or publicity and establish the facts. This includes reading beyond the headline or abstract, looking critically at graphical displays and numerical information, and checking that the data support the study results.

*Are all patients enrolled in the trial accounted for?*

Data on all subjects entering a trial should be analysed with respect to the groups to which they were initially randomised. In other words, the data on a trial participant who was randomised to intervention X should be analysed along with the data on all subjects who received that intervention, even if for some reason this particular subject ended up receiving the alternative intervention or no therapy at all. This is known as **intention-to-treat analysis**.[9] Some subjects may drop out of the study because of adverse effects. Ignoring these people could give an overestimate of the benefits of the intervention.[10] Intention-to-treat analyses (vs "per protocol" or "on medication" analysis) are considered to produce results that are more akin to real life situations, where not all patients comply with their treatment.

*How are the results expressed?*

The way that treatment effects are presented influences the way in which the results are interpreted.[11] Consider how the use of percentages rather than absolute values can affect how differences are perceived. For example, a 20% price reduction differs in its monetary worth depending on the item being sold – it could represent 20p off the price of a pastry or a £2000 saving on a car (which reduction is more valuable to you will depend on your circumstances). Generally, treatment effects are reported in absolute or relative terms. These are explained on page 4. Calculating numbers-needed-to-treat or numbers-needed-to-harm can be useful for converting proportional risks into terms that are more easily related to individual patients.[12,13]

*Do the results presented match the study aims?*

Researchers should concentrate on results for outcomes they initially defined as being of primary importance. Greater emphasis should not be given to **secondary outcomes**, **subgroup analyses,** or **interim analyses** that produced positive results but which the study was not designed to investigate.[14,15] Such results might identify areas for future research, but they can be misleading and should be viewed with caution when considering current practice.

## Understanding the numbers

Consider two controlled trials of a treatment carried out in two different study populations over two years (Figure 1).[13]

In trial 1, conducted in high-risk patients, the event rate in the control group is 40% (blue bar). The event rate in the group receiving the treatment is 30% (red bar).

The **Absolute Risk Reduction** (ARR) is the simple difference in the event rates.

For trial 1, ARR = 40% - 30% = 10%

The **Relative Risk Reduction** (RRR) is the difference in event rates expressed relative to "what you started with".

For trial 1, take the 10% difference between the two groups and express it relative to the rate in the control group:  RRR = $\frac{10\%}{40\%}$ = 0.25 or 25%

Figure 1 (adapted from Barrett et al[13])

A **Number Needed to Treat (NNT)** estimates the number of patients a clinician will treat with a therapy over a given period of time for one patient to benefit. An NNT is the inverse of the absolute risk reduction, or if the ARR is expressed as a percentage, this should be divided into 100:

For trial 1, NNT = $\frac{100}{10}$ = 10

Nine out of ten patients will receive this treatment over two years without benefiting.

A **Number Needed to Harm (NNH)** is reported where positive effects are not seen, i.e. a treatment has a harmful effect.

The relative risk reduction appears more impressive than the absolute risk reduction, and the difference between them becomes larger as the event rate in the control group becomes lower. It is usually the relative risk reduction that appears in abstracts, advertisements, and in the press.
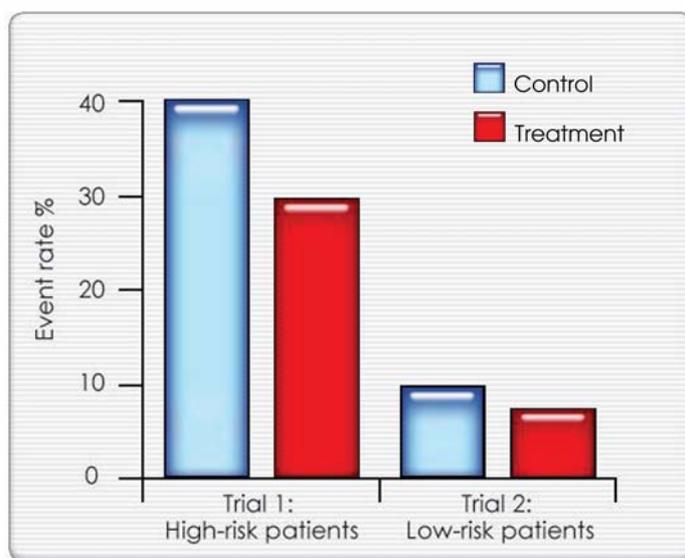
For trial 2, conducted in low-risk patients, the event rates are 10% in the control group (blue bar) and 7.5% in the treatment group (red bar).

The absolute risk reduction or ARR is

10% - 7.5% = 2.5%

The relative risk reduction or RRR is

$\frac{2.5\%}{10\%}$ = 25%

The NNT is $\frac{100}{2.5}$ = 40

Note that both trials show a relative risk reduction of 25%, but the absolute benefits were greater in the high-risk patients.  Trial 1 gives an NNT of 10 over two years whereas trial 2 gives an NNT of 40 over the same period.  This illustrates the importance of considering baseline risk before doing between-study or between-group comparisons.

*ARR = event rate in control group – event rate in treatment group*

*RRR = $\frac{event\ rate\ in\ control\ group\ –\ event\ rate\ in\ treatment\ group}{event\ rate\ in\ control\ group}$*

*NNT = 1/ARR   or   100/ARR (%)*

## More statistics…in brief

*"Of all the areas of mathematics, probability, and its inscrutable daughter statistics, are the most slippery to grasp. … I think that before any author drops any complex inferential statistics into a paper they should be obliged to give a commonsense interpretation of the data first:*

*"Eyeballing this data seems to show that as X increases so does Y. The statistics suggest this is unlikely to be a chance event that arose because we only sampled a part of the population…*

*There. That wasn't so hard."*

**Kevin Barraclough**, *general practitioner.*
BMJ 2004; 329: 1411.

## P values and confidence intervals

In comparative scientific studies, investigators test whether their observations are consistent with the assumption that two interventions have the same effect. An arbitrary probability of 1 in 20 or less ($P \leq 0.05$) is used to define a statistically significant difference between groups.

A $P$ value does not indicate the magnitude of a difference, i.e. the clinical significance of a result. Furthermore, a $P$ value does not give any indication as to the strength of the evidence obtained from a study.[16] The best estimate of this can be made by calculating confidence intervals for the results of statistical tests.

Calculating the confidence interval around a result gives you a range of values within which you can expect the true value for a population to fall with a specific level of certainty (usually 95%).

## Resources

NHS Wales e-library for health - http://howis.wales.nhs.uk - provides access to:

Bandolier – Evidence-based Healthcare

Clinical Evidence

Cochrane Library

Drug and Therapeutics Bulletin

full-text online journals including the BMJ

MEDLINE, EMBASE, & Evidence Based Medicine Reviews

National electronic Library for Health - www.nelh.nhs.uk - also provides links to:

Hitting the Headlines service from the Centre for Reviews and Dissemination

The National Institute for Health and Clinical Excellence (NICE)

PRODIGY Knowledge

Centre for evidence based medicine – www.cebm.net

The National Prescribing Centre - www.npc.co.uk

Using evidence to guide practice. MeReC Briefing and Supplement 2005; 30.

Medicines and Healthcare products Regulatory Agency – safety information from the Committee on Human Medicines - www.mhra.gov.uk

*Patient oriented materials*

Best Treatments website – clinical evidence for patients – www.besttreatments.co.uk

Treatment Notes – www.dtb.org.uk

Visual Rx – allows generation of visual tools for demonstrating risks – www.nntonline.net

Texts:

Greenhalgh T. How to read a paper. London: BMJ Publishing Group, 1997.

Guyatt G, Rebbie D eds. Users' guides to the medical literature. Chicago: American Medical Association Press, 2002.

Sackett D etal. Evidence-based medicine: how to practice and teach EBM. Edinburgh: Churchill Livingston, 1997.

Phillips CJ. Health Economics: an introduction for health professionals. Blackwell Publishing, 2005.

Paling J. Helping patients understand risks. 7 simple strategies for successful communication. Florida: Risk Communication Institute, 2006.

The WeMeReC education module on evaluating clinical evidence – case study and questions

*Are benefits reported alongside harms and costs?*

When assessing the balance of risks and benefits for an intervention, it is important to note that trial reports often present incomplete or inadequate safety data.[17,18] Randomised controlled trials are not usually sufficiently large or of sufficient duration to detect rare or long-term adverse effects. Information from postmarketing surveillance studies may need to be sought.

Finally, pharmacoeconomic data is not routinely presented in trial reports. Other research may need to be consulted to clarify cost implications. How study results are incorporated into clinical guidelines is often dependent on whether or not economic factors are considered alongside the clinical evidence (bodies such as The National Institute for Health and Clinical Excellence and the All Wales Medicines Strategy Group take economic factors into account when developing guidance).

*How does this affect my practice? What are the implications for my patients?*

It can be difficult to draw conclusions about how a study might impact your prescribing, particularly when many studies are designed to support licensing of a medicine, not its optimal use in general practice. Carefully selected study participants may not share the complexity and co-morbidity present in your patients, nor the potential to use a medicine for longer periods and, thus, the greater risk of unexpected adverse effects and drug interactions.

In the discussion section of a paper, authors should state whether, and to what extent, their study supported their hypothesis. They should also raise issues such as whether further research is needed.

*How do the results fit in with other information?*

The results of clinical trials need to be put into context.[2,19] The evidence from any study must be weighed alongside that from other relevant trials, with appropriate emphasis given to each according to its strength or validity. This is where reviews, systematic reviews, and meta-analyses can be valuable. Placing this information in the context of healthcare provision is how guidelines can help inform decision making. The same principles apply when evaluating these sources of evidence: they will only be as good as the methods they employ.

*Tip* You can gain the benefit of others' insight into the strengths and weaknesses of a study by consulting sources of summarised information, and by reading the correspondence that accompanies publication of a study. On-line sources may provide timely feedback, but bear in mind that reviewed comment can be more robust.

*In the end…*

Translating evidence on interventions into practice largely depends on how this information is conveyed to your patients. This is a subject in itself, and there are numerous valuable resources devoted to communication methods and aids. The *Resources* section on page 5 highlights some of these.

## References

1. Slawson DC et al. Becoming a medical information master: feeling good about not knowing everything. J Fam Pract 1994; 38: 505-513.
2. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" Lancet 2005; 365: 82-93.
3. Montori VM et al. Users' guide to detecting misleading claims in clinical research reports. BMJ 2004; 329: 1093-1096.
4. Greenhalgh T. How to read a paper. London: BMJ Publishing Group, 1997.
5. Schulz KF, Grimes DA. Allocation concealment in randomised trials: depending against deciphering. Lancet 2002; 359: 614-618.
6. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. Lancet 2005; 365: 1348-1353.
7. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. Lancet 2002; 359: 696-700.
8. Montori VM et al. Validity of composite end points in clinical trials. BMJ 2005; 330: 594-596.
9. Montori VM, Guyatt GH. Intention-to-treat principle. CMAJ 2001; 165: 1339-1341.
10. Fergusson D et al. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ 2002; 325: 652-654.
11. Framing. Bandolier. *http://www.jr2.ox.ac.uk/bandolier/ booth/glossary/framing.html*
12. Describing results of trials and reviews. Bandolier. http://www.jr2.ox.ac.uk/bandolier/booth/glossary/outputs. html
13. Barratt A et al. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. CMAJ 2004; 171: 353-358.
14. Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. Lancet 2005; 365: 1657-1661.
15. Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet 2005; 365: 176-186.
16. Sterne JAC, Davey Smith G. Sifting the evidence – what's wrong with the significance tests? BMJ 2001; 322: 226-231.
17. Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials – a systematic survey. BMC Clinical Pharmacology 2001; 1: 3. http://www.biomedcentral.com/1472-6904/1/3
18. Ioannidis JPA et al. Adverse events: the more you search, the more you find. (Editorial). Ann Intern Med 2006; 144: 298-300.
19. Sheldon TA et al. When to act on the evidence. Getting research findings into practice. BMJ 1998; 317: 139-142.