

Using evidence to guide practice

SUMMARY

The sheer volume of information being continually generated by numerous sources makes it very difficult for health care professionals to stay current with new developments and sift out the most useful information on interventions. This *Briefing* discusses approaches for sourcing the most useful evidence about interventions and explains some of the concepts necessary for its interpretation to help health care professionals and patients decide which interventions are appropriate for their care. It is supported by an **online supplement** that is available at www.npc.co.uk/merec.htm.

- Firstly, use trusted sources of pre-digested summaries of evidence or information (e.g. NICE guidance, systematic reviews in The Cochrane Library and reviews in Clinical Evidence) in preference to primary sources of information (see page 2).
- Consider the relevance of information obtained from different sources (see page 2):
 - Change of practice required?
 - Outcomes are patient-oriented rather than disease-oriented?
 - Frequency of the condition the information addresses?
 - Feasibility of the intervention discussed?
- Consider the factors that influence the validity of the information (see page 3), e.g.:
 - Study design
 - Comparator drugs and doses
 - Primary outcomes vs. secondary outcomes, sub-group analyses.
- Consider how the information is presented and interpreted (see page 5):
 - Statistical significance vs. clinical significance
 - Appropriate use of relative risks and absolute risks, NNTs and NNHs.
- Consider whether or not interventions are appropriate based on the information available and clinical expertise, alongside patient factors and preferences (see page 7). Comparisons of interventions should consider:
 - Safety
 - Tolerability
 - Effectiveness
 - Price
 - Simplicity.
- Consider how information can be relayed to patients to enable informed decisions to be made (see page 8):
 - Supplement descriptive terms with numbers
 - Use absolute numbers rather than relative risk alone
 - Express the odds of possible outcomes with a consistent denominator
 - Present the positive and negative perspective
 - Visual aids can be useful.

Background

There is an ever-increasing amount of information available to health care professionals on the evidence for different interventions, and from a range of sources. This presents a number of challenges to its use. The sheer volume of information being continually generated makes it very difficult for health care professionals to stay current with new developments.¹ In addition, not all available information is consistent, reliable and relevant.² Even once useful information is identified, it then has

to be accurately interpreted and effectively communicated to patients so that appropriate joint decisions about individual patient care can be made.

So what is useful information?

Health care professionals need information to keep up to date with new developments and to answer specific questions related to individual patients.^{1,2} There are various sources of information, including: original studies published in professional journals; text books; promotional

Panel 1: Usefulness of various sources of information

$$\text{Usefulness of information} = \frac{\text{relevance} \times \text{validity}}{\text{time to locate and interpret}}$$

Examples of usefulness equation applied to different information sources:

- Refined syntheses of information from trusted sources (e.g. NICE guidance, The Cochrane database of systematic reviews, Clinical Evidence):
– High validity and little time to locate and interpret. If relevant at that time then very useful.
- Individual primary research papers (e.g. identified via MEDLINE searches):
– Validity needs to be properly assessed and time is required to locate and interpret. May be relevant at that time, but usefulness decreases as effort increases.

Validated, refined syntheses of information, obtained from trusted sources, are likely to be more useful to busy health care professionals than primary sources of information that require careful, time-consuming evaluation

materials produced by manufacturers and delivered by sales representatives; nationally produced information and clinical guidelines; and colleagues or local experts. These sources all have their strengths and weaknesses, and none of them alone can meet all needs all of the time. However, some information can be more useful in certain contexts than others.

The usefulness of information may be considered in terms of three interacting attributes: its relevance to every day practice; its validity; and the amount of effort or work required to obtain it (see **Panel 1**).^{1,3} The greater the relevance and validity of the information, or the less the amount of work required to obtain it, the greater the usefulness of the information source. However, not all information that is easily obtained is highly relevant or valid. When considering or seeking evidence or information, the use of a filtering process will help to reduce the amount of time involved and improve efficiency. Firstly determine whether the information is relevant. If the information is found to be relevant, then consider validity.

Relevance

How relevant a piece of information is depends on how applicable the information is to everyday practice. National policy and guidance, such as that issued by the National Institute for

Health and Clinical Excellence (NICE) and the Department of Health, will be applicable across many areas of everyday practice. However, other sources of information may be less so. The COFF acronym outlines the general factors to consider when deciding how relevant a piece of information is:⁴

- Will the information require a **C**hange in current practice?
- Do the **O**utcomes described in the information demonstrate that patients will live longer and better lives? (i.e. are they patient-oriented outcomes rather than disease-oriented outcomes? See **Panel 2**)
- Is the condition discussed seen **F**requently?
- Is the intervention discussed **F**easible in local practice?

If the answer to any of these questions is 'no' then the information is not likely to be immediately relevant. These factors serve only as a general guide. There will be occasions where information relating to an infrequently encountered condition becomes more relevant, but it would be more practical to obtain it at the relevant time and prioritise information on the frequently encountered conditions.

Validity

Validity refers to the extent to which the information or evidence can, or should, be relied upon as being correct, and assessment of validity is the realm of critical appraisal. An appreciation of how the design and conduct of trials and other studies can bias their results is necessary for the practice of evidence-based medicine, and some basic concepts that health care professionals need to be aware of are discussed later. However, **critical appraisal is difficult to do well and is time consuming.**

The usefulness of information is influenced by the amount of work or effort required to obtain it. Therefore, it is not always appropriate or practical for busy health care professionals to critically appraise primary sources of information. Instead, **validated, refined syntheses of information, obtained from trusted sources, are likely to be more useful** to busy health care professionals. Guidance issued by NICE, systematic reviews available from The Cochrane Library and reviews in Clinical Evidence are amongst the most useful sources of information that are available on the evidence for particular interventions. They employ the most rigorous methods and make conclusions based on the most robust evidence, and they are easily accessed online. Once a source of information such as these is determined to be relevant, there is little need to consider its validity, as this has already been done. In contrast, individual, primary sources of information identified via MEDLINE searching may provide *an* answer to a specific question, but without consideration of the wider information

Panel 2: Patient-oriented evidence vs. disease-oriented evidence⁵

Patient-oriented outcomes (POOs)

matter to patients and help them live longer or better lives, for example, mortality, morbidity (e.g. non-fatal myocardial infarction, stroke), quality of life, clinical events (e.g. number of hospitalisations)

Disease-oriented outcomes (DOOs)

are intermediate, physiological or surrogate results, for example, blood sugar level, blood pressure, lipid levels, coronary plaque thickness

Examples of how consideration of DOOs can provide different information compared with POOs:

Doxazosin for hypertension:

- DOO: reduces blood pressure in black hypertensive patients
- POO: increases cardiovascular events compared with chlortalidone

HRT for prevention of cardiovascular events:

- DOO: reduces LDL-cholesterol and increases HDL-cholesterol levels
- POO: no decrease in cardiovascular or all-cause mortality, increase in cardiovascular events in women aged > 60 yrs with combined HRT

and evidence base, may not provide the *best* answer (see **Panel 1**).

What makes some sources of information or evidence more valid than others?

To describe all the influences on validity in all types of study is beyond the scope of this *Briefing*. An overview of validity in studies comparing one intervention against another, such as in randomised controlled trials (RCTs), is discussed below. This aims only to introduce some basic concepts and explain why some studies provide more valid evidence than others. There are several resources available for those interested in developing a deeper understanding of critical appraisal skills (see the **online supplement** at www.npc.co.uk/merec.htm). However, for most health care professionals this may not be practical.

What forms of evidence are available?

There are many types of study design and a description of the more common types and forms of evidence is given in the **online supplement**.

Studies that are conducted retrospectively (i.e. after an event of interest has occurred), as in case-control studies and some cohort studies, are susceptible to bias from a number of sources. The selection of cases and controls, the accurate recording of exposure to an intervention and the reporting of events of interest may all influence the outcome of the study and can be difficult to control for. In contrast, double-blind RCTs, in which subjects are truly randomly allocated to the study interventions and neither the investigators nor the subjects are aware of which intervention the subjects will receive, can control for some of the possible biases encountered with retrospective studies.⁶ Therefore, well conducted, prospective double-blind RCTs are considered more reliable (less open to bias) than unblinded prospective studies, which are considered more reliable than retrospective studies.

This creates a hierarchy of evidence that describes the validity of different sources of evidence for interventions (see **Panel 3**). In general, well-conducted systematic reviews and meta-analyses of robust, double-blind RCTs are considered the gold standard of evidence for making decisions about interventions. However, for information about diagnosis, prognosis and harms, other sources and data may be more appropriate.^{6,8} In addition, there may be occasions where no relevant RCT has been conducted and the next best available evidence will need to be considered.⁹

What factors can influence the validity of clinical studies?

Some of the more pertinent influences on study validity that health care professionals should be aware of are discussed below.

Panel 3: Hierarchy of strength of evidence used in grading recommendations in NICE clinical guidelines (strongest to weakest)⁷

Ia	evidence from systematic reviews or meta-analyses of randomised controlled trials
Ib	evidence from at least one randomised controlled trial
IIa	evidence from at least one controlled study without randomisation
IIb	evidence from at least one other type of quasi-experimental study
III	evidence from non-experimental descriptive studies, such as case-control studies
IV	evidence from expert committee reports or opinions or clinical experience of respected authorities

Study populations. Using evidence from clinical studies to inform treatment decisions involves accepting the 'average' results obtained in a group of people and generalising these to individual patients. Clinical studies often employ strict subject inclusion and exclusion criteria and, whilst this approach can help to remove some potential biases in the study, it may mean that the characteristics of subjects in the study do not reflect the characteristics of all patients who will receive the intervention.³

Possible reasons why study subjects may not reflect patients seen in routine clinical practice include different risk factors for disease, the presence of co-morbidities, and differing levels of care received in different care settings.^{3,6} These factors, and many potential others, can mean that responses or susceptibilities to the effects of interventions may differ between those observed in study subjects and those potentially seen in actual patients. This influences how applicable the results of some studies are to the broader real life patient population.

Comparator drugs and doses studied. When studies compare an intervention of interest against an active control, it is important that any comparisons that are made are fair. For example, comparison of a low dose of a drug of interest with a high dose of a control drug will have the effect of minimising the risk of side effects from the drug of interest and potentially maximising the harms seen with the control drug. Similarly, comparing a high dose of the drug of interest with a low dose of an active control drug will potentially maximise the efficacy of the drug of interest, but may minimise the efficacy of the control drug. Both of these approaches could potentially unfairly bias the studies in favour of the drug of interest.

Sample size and power. Clinical studies such as RCTs are designed to test whether a difference exists between two or more interventions in terms of specific outcomes or endpoints. For a study to detect a statistically significant difference between treatments, the study must be large enough (i.e. have enough subjects participating in the study) for a sufficient number of endpoints of interest to occur. The ability of the study to reliably detect a difference between interventions, if one exists, is sometimes referred to as the 'power' of the study.⁹

Example: The LIFE study was the first RCT to show that an angiotensin II receptor antagonist (losartan) can reduce cardiovascular morbidity and mortality in people with hypertension. However, LIFE was conducted in a select group of high risk patients who had left ventricular hypertrophy. These patients are atypical of the majority of hypertensive patients (see *MeReC Extra* Issue No. 5).

Example: The MELISSA study compared the short-term gastrointestinal (GI) tolerability of the COX-II selective inhibitor meloxicam at a dose of 7.5mg per day against the non-selective NSAID diclofenac at a dose of 100mg per day in patients with osteoarthritis. There were significantly fewer GI adverse events and withdrawals due to adverse events with meloxicam than diclofenac, but efficacy measures favoured diclofenac. This led to some suggestions that the doses used in the study were not comparable (see *MeReC Briefing* Issue No. 20).

Example: The LIFE study had a composite primary endpoint of cardiovascular death, myocardial infarction and stroke. Secondary endpoints included total mortality, hospitalisation for angina or heart failure, and the need for revascularisation. A prespecified sub-group analysis was performed in patients with diabetes (see *MeReC Extra* Issue No. 5).

A calculation of the number of subjects required for a study to have sufficient power must be made before a study begins. Clinical study reports should ideally indicate that a power calculation has been made and it is common for studies to stipulate a power of 80–90%.⁶ It is important to remember that the power of a study refers to its ability to detect a difference only in that endpoint on which the power calculation is based, i.e. the primary endpoint. Studies may not be sufficiently powered to reliably detect differences in other (secondary) endpoints or in subgroups.

Primary vs. secondary endpoints and sub-group analyses. Clinical studies such as RCTs are designed to answer specific questions, e.g. whether one drug is more effective than another at preventing a particular event occurring. The specific event that the study is designed to assess the effect of the drugs upon is called the primary endpoint. The sample size required to adequately power a study is dependent on the number of primary events that are expected to occur over a given time period. Depending on the characteristics of the population being studied, the incidence of some events of interest (e.g. death) may actually be quite low in the period over which the study can be conducted. This means that, in order to reliably study the effectiveness of two different interventions in preventing the event of interest, many study subjects would need to be enrolled in the study. This poses many difficulties, as such studies are very expensive to conduct and recruitment of sufficient numbers of subjects into the study can be difficult.

To overcome these difficulties, studies are frequently designed and powered to detect a difference in a composite of events (e.g. cardiovascular death and/or non-fatal myocardial infarction and/or any kind of stroke), rather than just one. Studies designed with such a composite primary endpoint (e.g. any major cardiovascular event) require fewer subjects, as the total number of events to be detected is in effect increased. However, when considering the results of studies that combine several event types into one composite endpoint, it can be difficult to determine the true effect of the intervention on each of the event types.¹⁰

Secondary endpoints are additional events of interest, but which the study is not specifically powered to assess. Therefore, analyses of secondary endpoints need to be viewed with caution. Similarly, sub-group analyses, in which primary or secondary endpoints are considered in a smaller sub-group of the whole study population (e.g. only those subjects with a specific baseline characteristic) need to be considered with care, as the power of the study is based on the entire study population.

Secondary endpoints and any sub-group analyses should be pre-specified in the study

protocol and methods section of the study report. If not, there is a possibility that the data from the study has been analysed in many ways to find any kind of positive result in favour of the intervention. Such 'data dredging' cannot be relied upon alone as providing reliable information. This is because if a set of data is analysed and re-analysed enough times, a positive result may occur purely by chance.¹¹

Study duration. As there is a need for enough events of interest to occur for a clinical study to be adequately powered, the duration of the study needs to be sufficient to allow those events to occur in the given sample of subjects. If study follow-up is not long enough, too few events may occur to enable detection of any difference between interventions. In addition to this, information about interventions that could ultimately be used chronically in patient populations may only be initially available from studies of limited duration. The results of clinical studies of limited duration cannot provide information on the longer-term effects of treatment. For example, if a clinical study showing benefit for an intervention over a six-month period was continued to 12 months, that benefit may no longer be apparent. Also, clinical studies of limited duration cannot provide information about relatively rare adverse effects.¹²

On occasions, a clinical study will be terminated earlier than planned because one of the interventions has demonstrated significant benefits or ill-effects compared with its comparator. It could be considered unethical to continue to withhold treatment with an intervention from those people who have been treated with an inferior comparator in the study. However, early termination of the study limits the information available on longer-term effectiveness and side effects of the intervention of interest.

Example: ALLHAT remains the largest hypertension trial to date. Patients were randomised to chlortalidone, amlodipine, lisinopril or doxazosin as initial antihypertensive therapy for four to eight years of follow-up. However, the doxazosin arm was stopped prematurely because of a higher rate of stroke and heart failure compared with chlortalidone (see *MeReC Briefing* Issue No. 29).

Newly marketed medicines are designated as black triangle (▼) drugs under intensive safety surveillance. This is because, when a new medicine is marketed, only limited information on adverse effects is available from studies involving relatively few subjects and conducted for relatively short periods of time.¹² Such medicines should be used with greater caution and, in general, it would seem sensible to avoid their routine, wide-scale use until a substantial body of evidence of safety (and effectiveness) is accumulated from a wider population than initial clinical studies can provide.

Follow up. The rate of follow up (the proportion of subjects originally entering the study for whom data are available at the end of the study) is important when considering the study results. If data are available on too few original subjects (e.g. <80%) the power of the study will be diminished and the results will be unreliable.^{6,8}

Intention to treat analysis. Subjects may prematurely withdraw or drop out of a clinical study for a number of reasons that may be related to the study intervention (e.g. adverse effects of the interventions) or unrelated (e.g. moves away). It is important that all subjects who enter a study are accounted for when results are analysed, as failure to account for all subjects can bias the study results. For example, if some subjects assigned to the intervention of interest were withdrawn from a study due to adverse effects, this would reduce the chance of them experiencing an event of interest.¹ If they were not included in the final analysis, this could artificially inflate the response rate for the intervention. To overcome this, all subjects who are assigned to an intervention at the start of the study should be included in the analysis of that intervention group, whether or not they completed the study or received the intervention. This is called 'intention to treat' analysis and more closely represents the real life situation where some patients are not compliant with interventions.^{6,8,13}

Sponsorship and publication bias. Systematic reviews of clinical studies have found that those sponsored by drug companies were more likely to have outcomes favouring the sponsor than were studies funded by other sources.^{14,15} This is not because studies sponsored by the industry were of poor quality,¹⁴ but may be related to the selection of an inappropriate comparator in the study,¹⁴ or biased interpretation of results.¹⁶

In addition, industry sponsorship has been associated with restrictions on publication and data sharing, i.e. publication bias.¹⁴⁻¹⁶ There are many possible reasons for this, including journal editorial boards favouring publication of studies with positive results rather than neutral or negative studies.¹⁷ However, this may mean that only studies that are favourable to the sponsor are made widely available, and less favourable studies have a more restricted availability. Studies with positive findings can also sometimes lead to multiple publications,³ which may create the impression that more evidence exists in favour of a drug than is actually the case. A global clinical trials register that will disclose all trials and results, regardless of outcome, has been proposed to alleviate some of these issues.¹⁷

How are results presented in studies?

Clinical studies are replete with statistics and other numerical expressions that can appear

complicated and intimidating to those who are unfamiliar with how they are used and their meaning. However, it is not necessary to be a statistician to understand and interpret the more common methods of expressing study results. A brief overview of these methods is provided below, using the results of the CURE study¹⁸ for illustrative purposes (see **Panel 4**, page 6). The **online supplement** provides more details of how the various ways of presenting results are derived.

What are p-values?

Statistical tests cannot absolutely prove anything; all they can do is quantify the likelihood that an observed result in a study is a real effect rather than due to chance. Tests of significance (hypothesis tests) in clinical studies are carried out to assess the probability that an observed difference between interventions could have occurred by chance. However, rather than being tests to check if a difference exists between the effects of different interventions, they are actually tests to check the hypothesis that no difference exists between interventions (referred to as a 'null hypothesis').

In simple terms, the p-value is the probability that no difference exists between interventions for a given endpoint. As probability can take any value between zero (no chance at all) and 1.0 (certainty), so can the p-value. The closer the p-value is to zero, the less chance there is that the effects of two interventions are the same. By an arbitrary convention, if the p-value is ≤ 0.05 (which means that the probability of the effects of two interventions being the same is 1 in 20 or less) the effects of two interventions are said to be statistically significantly different. Conversely, if the p-value is >0.05 , this, by convention, would indicate there is no statistically significant difference in effect between the interventions.

This approach has some drawbacks. For example, $p=0.049$ would be considered statistically significant, but $p=0.051$ would not, despite there being very little difference in these p-values. In addition, significance tests alone do not indicate the magnitude of the observed difference between treatments that is needed to determine the clinical significance of study results (see below).

What are confidence intervals?

Confidence intervals (CI) provide more useful information than the p-value alone can. The CI around a result obtained from a study sample indicates the range of values within which there is a specific level of certainty (usually 95%) that the true population value for that result lies.

If a study finds that the CI around the difference in mean effects of two interventions contains the value zero, then we cannot rule out the possibility that there is no difference in effect between the interventions; the difference

Example: The p-value for the difference between major bleeding rates in the CURE study = 0.001. This means the probability that no difference exists between the bleeding rates is 1 in 1000. This is less than the conventional probability of 1 in 20, so this would suggest that there is a statistically significant difference in major bleeding rates.

Panel 4: Summary results of the CURE study, to be used for illustrative purposes.¹⁸

CURE was a RCT of a combination of clopidogrel and aspirin vs. aspirin alone in patients with acute coronary syndromes that was published in 2001 (see *MeReC Bulletin* Vol 15 No. 6). After an average of nine months the composite endpoint of cardiovascular death, non-fatal myocardial infarction or stroke was lower in patients taking clopidogrel plus aspirin than in patients taking aspirin alone (9.3% vs. 11.4%, relative risk 0.80 [95% CI 0.72–0.90]). As expected from the additive antiplatelet effects, there were more major bleeds in the group taking the combination compared with aspirin alone (3.7% vs. 2.7%, $P=0.001$).

Example: The 95% confidence interval around the primary endpoint relative risk estimate in the CURE study was 0.72 to 0.90. Therefore, we can be 95% sure that the true relative risk lies within the range 0.72 to 0.90. As relative risk is a ratio and we are 95% sure the relative risk is between 0.72 and 0.90 (i.e. does not include the value 1.0), this would suggest that there is a statistically significant risk reduction with combination therapy compared with aspirin alone.

between the interventions would not be statistically significant. However, if the CI excludes the value zero, we can be reasonably (95%) certain that there is a difference between the interventions; the difference would be statistically significant.

When a CI is constructed around a statistic that is a ratio (e.g. relative risk, odds ratio, hazard ratio — see below), if the CI does not contain the value 1.0, this would indicate that a statistically significant difference exists. However, if the CI contains the value 1.0, then this would indicate no statistically significant difference.

CIs can also be constructed around the measure of effect of each separate intervention. If the CIs around each measure of effect do not overlap, this would indicate that the effects of each separate intervention are significantly different. However, if they do overlap it is less certain whether there is or is not a significant difference between the effects of each intervention. The width of the CI (i.e. the range of values of the CI) can also provide useful information. A CI that is tight around the point estimate indicates that the study has sufficient power to be relatively precise, but if the CI is very wide, this indicates that the study could be underpowered and the point estimate imprecise. See the **online supplement** for further explanation of the interpretation of CIs.

Statistical vs. clinical significance?

Just because a study finds a statistically significant difference in effect between interventions does not mean that the results are clinically significant. Studies involving large numbers of subjects can find statistically significant differences between interventions that actually represent very small effects. Therefore, it is important to consider whether differences between interventions that are labelled statistically significant are clinically worthwhile.

What are absolute risks, relative risks, odds ratios and hazard ratios?

The method of presenting the results of clinical studies can affect their interpretation by clinicians¹⁹ and non-clinicians alike.^{20,21} Therefore, it is important to understand the different ways in which results can be presented.

Absolute risk refers to the simple event rate in a group of people who receive an intervention.

Example: In the CURE study, the absolute risk of the primary endpoint in the clopidogrel plus aspirin (combination) therapy group was 9.3% and in the aspirin alone group was 11.4%.

Relative risk (RR) estimates the size of effect of an intervention of interest relative to the size of effect of a comparator.

Example: In the CURE study, the crude RR of the primary endpoint in the combination therapy group compared with the aspirin alone group would be calculated as $9.3\%/11.4\% = 0.82$. (NB the RR in the study was calculated as 0.80 using cumulative rates over time — see Kaplan-Meier curves below).

Absolute risk reduction (ARR) is the difference in event rates between two interventions. To be most useful, ARR must be set in the context of the underlying incidence of the event of interest. Without this information about the underlying incidence and risk of the event of interest, we would not know if an ARR of 1% represented a change in risk from 2% to 1% or from 21% to 20%.

Example: In the CURE study, the ARR in the primary endpoint with combination therapy compared with aspirin alone would be calculated as $11.4\% - 9.3\% = 2.1\%$.

Relative risk reduction (RRR) is the reduction in risk of an event brought about by one intervention relative to the risk of the event in people using a comparator intervention. Without further information about the underlying incidence and risk of the event in the population, we would not know whether this relative reduction in risk represents a worthwhile benefit. Results presented as RRR may appear more impressive than results presented as ARR. Therefore, RRR needs to be set in the context of the underlying incidence of the event to be meaningful.²²

Example: When CURE was published in 2001, headlines claimed that the combination of clopidogrel plus aspirin reduced the risk of heart attack, stroke and CV death by 20%. This is the RRR, calculated as $1-RR$ (i.e. $1-0.80 = 0.20$). This may appear to some as being more impressive than the 2.1% ARR it actually represents.

Odds ratio (OR) expresses the odds of having an event compared with not having an event in two different groups. As long as the risk of the event of interest is low, OR and RR are approximately equal. This is the case in most RCTs, so the use of one measure or the other is unlikely to have an important influence on treatment decisions. However, as the risk of the event of interest increases, estimates of RR and the OR diverge,¹¹ so OR and RR should not be treated as the same in studies of subjects at high risk of events. For this reason, OR rather than RR should be used to express results in case-control studies.

Number needed to treat (NNT) expresses the number of people who would need to receive an intervention to prevent one event of interest. It is calculated by taking the reciprocal of ARR. The smaller the NNT, the greater the effectiveness of the intervention in the study population. NNT is a useful, intuitive way of representing study results, but it must be remembered that, for making comparisons between different interventions by using their NNTs, like must be compared with like. An NNT for one intervention to prevent an event that is calculated on data collected over nine months can only be compared directly with an NNT for another intervention if it relates to the same type of event calculated on data collected from a similar population over the same length of time.

Example: In the CURE study, the ARR of the primary endpoint with combination therapy compared with aspirin alone was 2.1% over nine months. The NNT would be calculated as $1/ARR = 1/0.021 = 48$. Therefore, 48 people would need to be treated with combination therapy instead of aspirin for nine months to prevent one primary endpoint occurring.

Number needed to harm (NNH) expresses the number of people who would need to receive an intervention for one person to suffer a harmful event of interest. As with NNTs, like must be compared with like when considering NNHs for different interventions. The larger the NNH, the less harmful the intervention with respect to the harmful event considered.

Example: In the CURE study, the absolute increase in risk (ARI) of a major bleed with combination therapy compared with aspirin alone is calculated as $3.7\% - 2.7\% = 1\%$. The NNH would be calculated as $1/ARI = 1/0.01 = 100$. Therefore, for every 100 people treated for nine months with combination therapy instead of aspirin alone, one will suffer an additional major bleed.

Kaplan-Meier survival curves (K-M curves) are increasingly used to represent outcomes that are times to an event. Although it is conventional to discuss analyses of such events as survival analyses, these may also be positive events, such as discharge from hospital.²³ In prospective clinical studies, such as RCTs, subjects are recruited over a given period and are followed up to a fixed date. Therefore, the last subjects recruited will be observed for a shorter period and will be less likely to experience an event of interest than the first subjects recruited. In addition, subjects may be lost to follow-up during the study for unrelated reasons. For these subjects, survival time is said to be censored, and special techniques are used to account for them in the analyses.^{23,24}

K-M curves represent the proportion of the study population still surviving (or free of disease or some other outcome) at successive times.¹¹ Curves for the intervention of interest and the comparator are often represented on the same graph and a p-value can be calculated to determine the likelihood that there is no

difference between the two survival curves. As the number of subjects in each intervention group decreases over time, the curves are more precise in the earlier periods (left hand side of the survival curves) than later periods (right hand side of the survival curves).¹¹ To account for this, the RR of the event of interest over the entire study period needs to be weighted for the number of subjects available over time. This weighted RR is called the **Hazard Ratio**.

Understanding diagnostic tests

It may be thought that diagnostic tests provide conclusive information about the presence or absence of a particular disease state or condition. However, this is not so.²⁵ Although test results may be reported simply as 'positive' or 'negative', the reliability of these results can be influenced by how common the disease state or condition is in the population being tested and a number of properties of the tests used. It is therefore necessary to understand how a particular test result predicts the risk of a disease or condition in an individual in clinical practice.²⁶ The **online supplement** explains the concepts of accuracy, sensitivity and specificity, predictive values and likelihood ratios that are necessary to appreciate the limitations of diagnostic test results.

Decision-making and communication

Having sourced or received a relevant piece of information and considered the factors that influence its validity, that information needs to be incorporated into the decision-making processes involved in patient care. At some point this will involve making a decision on whether or not an intervention is appropriate, based on that information, the clinical expertise of the health care professional, and patient factors and preferences.

Deciding whether one intervention is more appropriate than another can be a challenge. The STEPS framework may be useful to help build an overall picture of the benefits and harms of one intervention over another.

What is STEPS?

The STEPS acronym outlines the factors that are useful to consider when comparing one intervention with another (see **Panel 5**, page 8).³ Patient factors and preferences are inherent considerations in the STEPS framework, and it is important that patients are involved in decision-making processes as fully as possible. Explaining risks to patients in an effective way is essential in ensuring that their decisions about treatment are informed.^{26,27}

How is risk best explained?

The values and perceptions of individual patients, and their attitudes to risk, may be different to those of health care professionals.²⁸

Panel 5: The STEPS acronym for comparing interventions³

Safety. Remember that only limited safety information is available from prospective clinical studies like RCTs. Safety data from real-life comparison studies may be more useful as these will better reflect the patient population.

Tolerability. Rates of withdrawal of interventions in clinical studies can be an indicator of whether interventions have effects that will influence compliance.

Effectiveness. Does the drug work (in the real world)? Head-to-head comparisons of interventions will help answer this question. If information comes from different trials, care must be taken to ensure that comparisons are valid (e.g. study populations, drug doses used, etc.).

Price. All costs should be considered. For example, costs of visits for drug titration or monitoring, in addition to the purchase cost of the drug.

Simplicity of use. Factors such as the frequency of administration or ease of use of a delivery device could be important for patients and make the difference between compliance and non-compliance with therapy.

Interventions to reduce the risk of an event do not completely eliminate that risk

There are many influences on individual perceptions of medical interventions, including previous experience, media reporting and culture.²⁸ Just as the way that data is provided to clinicians can influence their interpretation and perceptions,¹⁹ the way that information is provided to patients can also influence their perceptions,^{21,29} which may influence their decisions to accept or refuse an intervention. Therefore, it is necessary to have an appreciation of the different ways in which risk can usefully be explained.

It is important that patients understand that interventions to reduce the risk of an event do not completely eliminate that risk. There will be some patients who, due to their use of the intervention, will not experience the event.

However, there will be others who experience an event, even though they are also using the intervention, and there is no way of knowing which patients will benefit and which will not. It is also important that patients understand the harms associated with the use of interventions.

Patients' abilities to assimilate information vary and it is important not to overload them. When conveying information, the use of descriptive terms such as "low risk" can be interpreted differently by patients and health care professionals. Therefore, explanations of risk should be supported by numbers:²⁶

- Use absolute numbers rather than relative risk alone. For example, a 25% reduction in the risk of dying of breast cancer sounds more impressive compared with the absolute risk reduction of 1 in 1000 women it could actually represent.²⁶
- Express the odds of possible outcomes with a consistent denominator. For example, 1 in 1000 experience a side effect and 5 in 1000 experience a beneficial effect, rather than 1 in 1000 experience a side effect and 1 in 200 have a beneficial effect.
- Present the positive and negative perspective around possible outcomes. For example, if the chance of experiencing a side effect is 1 in 1000, the chance of not experiencing a side effect is 999 in 1000.
- Visual aids can be useful to help describe probabilities of events occurring or not (see the Resources section of the **online supplement**).^{26,30}

References

- 1 Slawson DC, Shaughnessy AF, Bennet JH. Becoming a medical information master: feeling good about not knowing everything. *J Fam Pract* 1994;38:505-13.
- 2 Grandage KK, Slawson DC, Shaughnessy AF. When less is more: a practical approach to searching for evidence-based answers. *J Med Libr Assoc* 2002;90:298-304.
- 3 Peglar S, Underhill J. Evaluating promotional material from industry: an evidence-based approach. *Pharm J* 2005;274:271-4.
- 4 Maskrey N, Peglar S, Underhill J. Working smarter with information [Letter]. *Rapid response to BMJ* 2005;331:352. Available from: <http://bmj.bmjournals.com/cgi/eletters/331/7512/352-b>. Accessed 31/08/05.
- 5 Ebell MH, Siweck J, Weiss BD, et al. Strength of recommendation taxonomy (SORT): a patient-centred approach to grading evidence in the medical literature. *Am Fam physician* 2004;69:548-56.
- 6 Greenhalgh T. How to read a paper. London: BMJ Publishing Group 1997.
- 7 National Institute for Clinical Excellence. Chronic obstructive pulmonary disease. Clinical Guideline 12; February 2004. Available from: www.nice.org.uk. Accessed 22/09/05.
- 8 Sackett DL, Richardson WS, Rosenberg W, et al. Evidence-based medicine: how to practice and teach EBM. Edinburgh: Churchill Livingstone 1997.
- 9 Sackett DL, Richardson WS, Rosenberg W, et al. Evidence-based medicine: what it is and what it isn't [Editorial]. *BMJ* 1996;312:71-2.
- 10 Montori VM, Permyer-Miralda G, Ferreira-González I, et al. Validity of composite end points in clinical trials. *BMJ* 2005;330:594-6.
- 11 Guyatt G, Rennie D, eds. Users' guides to the medical literature. Chicago: American Medical Association Press 2002.
- 12 MHRA/CSM. Monitoring the safety and quality of medicines: New drugs under intensive surveillance (▼ list). Available from: <http://medicines.mhra.gov.uk>. Accessed 23/06/05.
- 13 National Prescribing Centre. Evidence based medicine. *MeReC Bulletin* 1995;6:45-8.
- 14 Lexchin J, Bero LA, Djulbegovic B, et al. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-76.
- 15 Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 2003;289:454-65.
- 16 Als-Nelson B, Chen W, Gluud C, et al. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921-8.
- 17 House of Commons Health Committee. The influence of the pharmaceutical industry. Fourth report of session 2004-5, Volume I. London: The Stationery Office Limited April 2005. Available from: www.parliament.the-stationery-office.co.uk/pa/cm200405/cmselect/cmhealth/42/42.pdf. Accessed 31/08/05.
- 18 The clopidogrel in unstable angina to prevent recurrent events trial investigators. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N Engl J Med* 2001;345:494-502.
- 19 Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *BMJ* 1994;309:761-4.
- 20 Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. *BMJ* 1995;311:1056-9.
- 21 Skolbekken J-A. Communicating the risk reduction achieved by cholesterol reducing drugs. *BMJ* 1998;316:1956-8.
- 22 National Prescribing Centre. An introduction to assessing medical literature. *MeReC Briefing* 1995;9:1-8.
- 23 Altman DG, Bland JM. Time to event (survival) data. *BMJ* 1998;317:468-9.
- 24 Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ* 1998;317:1572.
- 25 Greenhalgh T. Papers that report diagnostic or screening tests. *BMJ* 1997;317:540-3.
- 26 Paling J. Strategies to help patients understand risks. *BMJ* 2003;327:745-8.
- 27 Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ* 2003;327:741-4.
- 28 Thornton H. Patients' understanding of risk. *BMJ* 2003;327:693-4.
- 29 Freeman AC, Sweeney K. Why general practitioners do not implement evidence: qualitative study. *BMJ* 2001;323:1100-2.
- 30 Edwards A, Elwyn G, Mulley A. Explaining risks: turning numerical data into meaningful pictures. *BMJ* 2002;324:827-30.

The National Institute for Health and Clinical Excellence (NICE) is associated with MeReC Publications published by the NPC through a funding contract. This arrangement provides NICE with the ability to secure value for money in the use of NHS funds invested in its work and enables it to influence topic selection, methodology and dissemination practice. NICE considers the work of this organisation to be of value to the NHS in England and Wales and recommends that it be used to inform decisions on service organisation and delivery. This publication represents the views of the authors and not necessarily those of the Institute.