evaluation of studies of treatment or prevention interventions

In the first article in this series, we discussed how critical appraisal is an important step in evidence-based health care because some published healthcare research is too poor in quality to be safely applied to clinical practice. Critical appraisal is made easier by the availability of quality checklists, which can be used to appraise research studies systematically and efficiently. With practice, readers may no longer even need a checklist and should be able to decide whether an article is worth reading in a matter of moments.

Whether your clinical question is one of treatment, diagnosis, prognosis, or causation, there are 3 fundamental questions you should apply in deciding whether the research might help us to provide better care to our patients.1

Are the results of the study valid?
This question considers whether the results reported in the study are likely to reflect the true size and direction of treatment effect. Was the research conducted in such a way as to minimise bias and lead to accurate findings, or was it designed, conducted, or analysed in such a way as to increase the chances of an incorrect conclusion?

What were the results?
Once you have determined that the results are valid, it is important to gain an understanding of what the results really mean. If the new treatment is shown to be effective, how large is the effect? Is the effect clinically important? How precise is the treatment effect another way of asking how likely it is that the effect is real and not a result of the play of chance? The precision of a result is related to whether the study involved large numbers of people (which increases precision) or small numbers (which reduces precision).

Will the results help me in caring for my patients?
There are 2 concepts underlying this question. Firstly, you have to decide if the patients participating in the study are sufficiently similar to your patients, or whether there is a good reason why it would be inappropriate to apply the results to your patients. Secondly, are there risks or harms associated with the treatment, which might outweigh the benefits?

In this series of articles, we will use this framework to critique studies that address different types of clinical questions. In this EBN users’ guide, we will critique a study that evaluates an intervention and will confine ourselves to examining the first question on validity; in the next issue’s users’ guide, we will look closely at the second and third questions on results and their applicability.

Beginning with a clinical scenario, we will outline a search to identify high quality, relevant studies and will critique one of the studies using a series of intermediate questions (box) that together answer the fundamental questions above. We have used the original Users’ Guides to Medical Literature as a basis for our nursing series, and we will use the same approach of labelling the most important appraisal questions as “primary” and other appraisal questions, which really address the finer points of validity, as “secondary.”

Primary questions
- Was the assignment of patients to treatments randomised, and was the randomisation concealed?
- Was follow up sufficiently long and complete?
- Were patients analysed in the groups to which they were initially randomised?

Secondary questions
- Were patients, clinicians, outcome assessors, and data analysts unaware of (blinded to or masked from) patient allocation?
- Were participants in each group treated equally, except for the intervention being evaluated?
- Were the groups similar at the start of the trial?

Are the results of the study valid?
Clinical scenario
You are a diabetes specialist nurse who runs a foot care clinic with podiatrist colleagues. A patient presents at the clinic with a full thickness plantar foot ulcer, in the presence of signs of peripheral neuropathy but no evidence of significant arterial disease. The patient has read about artificial skin replacements on the internet and believes that these bio-engineered tissues will help her foot ulcer to heal much more quickly than conventional wound dressings. You are not sure and arrange to see her in a week, during which time you plan to search for research evidence on this intervention.

The search
You will recall that questions of the effectiveness of preventive and therapeutic interventions are best answered by randomised controlled trials (RCTs), and best of all by systematic reviews of RCTs. A search of the Cochrane Library does not identify any systematic reviews of skin replacements for diabetic foot ulcers. A search of Medline (Silver Platter) using the following search

(1) explode “Skin-Artificial” / all subheadings
(2) diabetic foot ulcer*
(3) 1 AND 2
(4) 3 AND (Publication Type = “Randomized-Controlled-Trial”)

identifies 1 article, which is a randomised trial comparing tissue engineered skin with traditional dressings for diabetic foot ulcers. You obtain the article from the library and sit down with your checklist.

Are the results of the study valid?
Primary questions
- Was the assignment of patients to treatments randomised, and was the randomisation concealed?

The purpose of randomisation is to create groups that are similar in all respects except exposure to the intervention. Through random assignment of patients to study groups, known and unknown factors (eg, age, sex, and disease severity) that could influence the outcome of the study are evenly distributed among groups. Methods of randomisation vary (eg, use of computer
generated numbers, tables of random numbers, and coin flipping); the important thing is that the method used ensures that all study participants have an equal chance of being assigned to each of the study groups. Methods of allocating participants alternately, or according to date of birth (odd or even years) or hospital record numbers do not give each participant the same chance to be included in each of the groups and should not be regarded as true randomisation methods. How can readers judge whether a study was randomised? The best way is to review the methods section of the paper for a description of how the randomisation was done and then determine whether the method ensured that patients had an equal chance of being in each group.

What should you do if there are no relevant randomised trials of the intervention? The next best study is an observational study, in which randomisation is not used to assemble the study groups. This design can be prone to selection bias because the investigators (who are likely to want the intervention to be effective) have control over who goes into each group and might choose the intervention group participants on the basis of those most likely to experience a positive outcome. Comparison of the results of randomised studies with observational studies has shown that observational studies almost always show larger, inflated, positive treatment effects.

The clinician (e.g., nurse, doctor, or other health professional) recruiting patients to a study should be unaware of the treatment group to which the next patient will be allocated. This is called allocation concealment. Analysis of previous studies has shown that if recruiters know the allocation schedule in advance, this may (consciously or unconsciously) influence their recruitment behaviour. For example, they might not recruit a patient who would be allocated to the control group or alter the sequence of recruitment so that patients who are more severely ill receive the new treatment. Such actions are often taken in the belief (in the absence of research evidence) that the new treatment is better than either the alternative (old) treatment or no treatment at all. Examples of strategies to conceal allocation include calling a central, coordinating office for each patient assignment; using numbered, opaque, sealed envelopes; and numbered or coded bottles or containers.

- **Was follow up sufficiently long and complete?**

This appraisal question has 2 components: the first refers to how long the patients are followed up, in order to see what happens to them as a result of their treatment. A certain amount of judgment is required when deciding whether duration of follow up is sufficient, but clinical practitioners are usually the people to judge. For example, if the trial is evaluating the effect of an intervention for a chronic health problem, or prevention of a health problem, the follow up phase of the study must be long enough to detect a clinically important effect, if it exists. Short follow up durations are unhelpful because the study may fail to capture sufficient numbers of patients who achieve meaningful outcomes.

It may seem obvious that every patient who is recruited to a trial should be accounted for at the end, but it seldom happens. If large numbers of participants are described as "lost to follow up," it throws doubt on the validity of the results. Patients drop out of studies for non-random reasons. In this sort of situation we do not know what happened to the "lost" patients, and they may have fared very differently from the patients who remained in the study. It is always possible that loss to follow up is caused by the intervention itself. For example, patients may discontinue their intervention due to side effects or because the medication made them feel worse. If patients who discontinued their medication were omitted from the analysis, the study would be left with only the patients who were more likely to be compliant and who had better outcomes. The end result would be that the medication would look better than it really is. Readers should check that the numbers presented in the analysis are close to the numbers initially randomised.

**SECONDARY QUESTIONS**

- **Were patients, clinicians, outcome assessors, and data analysts unaware of (blinded to or masked from) patient allocation?**

There are always 2 and sometimes 4 groups of people involved in a clinical trial, all of whom may be biased if they know which patients were allocated to which treatment. Studies are often labelled as single, double, or triple blind, depending on how many of these groups of people were unaware of the treatment allocation. Researchers should clearly state which groups were blinded and what steps were taken to minimise bias.

If patients know the group to which they have been allocated, they may have a heightened sensitivity to the good (or bad) effects of the treatment. In drug trials, patient awareness is usually avoided through use of a placebo that seems identical to the active treatment. It is often difficult or impossible to blind patients to nursing interventions, particularly if they have a psychosocial component. If clinicians caring for patients know the allocation, they may unwittingly alter the way they give other forms of care, and may also have a heightened sensitivity to good outcomes or adverse events in a way that biases the evaluation.

Ideally, the people who measure the outcomes are not the people who provide usual clinical care. If outcome assessors are aware of which group a patient is in, their measurement of key variables may be influenced. Distortion of outcome measurement (which is often unconscious) is less likely if the measure is an objective one, such as cotinine concentration as a marker of smoking status.

Although it is not always possible for patients, clinicians, and outcome assessors to be blinded to treatment group, it should...
always be possible for the data analysis to be done using coded data with no identification of treatment groups.

Readers of randomised trials should therefore look for evidence that patients, clinicians, outcome assessors, and data analysts were blinded to patient allocation. Wherever blinding was not possible, researchers should outline the steps they took to minimise bias.

- **Were participants in each group treated equally, except for the intervention being evaluated?**
  Because randomisation should ensure that the only systematic difference between study groups is the treatment in question, it is important that this principle is not undermined by extra care given to one group and not another (known as “co-interventions”). Clearly, if patients get an intervention plus some extras such as closer follow up or more time with a specialist nurse, it will be impossible to attribute any effects to the intervention itself. If clinicians are unaware of the allocation, then they will not deliver co-interventions to one group and not another. Readers of randomised trials should look carefully at the descriptions of the interventions received by all groups, particularly where clinicians are not blinded to allocation.

- **Were the groups similar at the start of the trial?**
  The process of randomisation should ensure that the groups are sufficiently similar at the start; researchers, however, should reassure themselves and their readers of this by presenting the baseline (or entry) characteristics of participants in each group. The characteristics described should be those that are known to, or may have an influence on the outcome of interest. If the trial has a small sample size, randomisation may fail to ensure that some factors are evenly distributed. The researchers may have done statistical tests to see if there were any significant differences in the baseline characteristics; more important than the significance of any difference is its size and whether the imbalance is likely to have affected the validity of the result. Imbalances in baseline characteristics that exist after randomisation can be adjusted for using statistical techniques, and readers should look for evidence of this. Readers can be most confident when results are consistent for analyses done with and without adjustment.

The 3 primary and 3 secondary appraisal questions outlined above can be applied to any study that aims to evaluate the effects of a preventive or therapeutic intervention, and will help readers to decide whether the results of a study are likely to be valid (ie, give a true estimate of the effect of the intervention). If you conclude that a study is valid, you should then consider the size of the effect: is the effect of sufficient clinical significance for you to want to use the intervention? In which patients? We will consider these questions in the next users’ guide.

**Answering the original question**

The randomised trial of tissue engineered skin compared with traditional treatment of diabetic foot ulcers met many, but not all, of the appraisal criteria. The study is described by the authors as a single blind, randomised controlled trial, although they do not say how randomisation was achieved, so we cannot judge the adequacy of concealment of allocation. Data are presented for 84% of patients at 12 weeks of follow up, although the authors state that follow up continued until 32 weeks. Only patients were masked to treatment in this study. Clinicians, outcome assessors, and data analysts could have been influenced in their management, measurement, and analysis by knowledge of which treatment each patient was receiving.

The groups appear to have been treated similarly. We are unable to determine if randomisation successfully distributed patients with particular baseline characteristics between the groups because the baseline data were not presented. We will be considering the results and applicability of this paper in the next users’ guide.

**NICKY CULLUM, RN, PhD**
Centre for Evidence Based Nursing, Department of Health Studies
University of York, York, UK