Clinically useful measures of the effects of treatment

In the EBN notebooks that have appeared in the previous 2 issues of the journal, we outlined 3 steps to help us to determine whether to apply the results of a research study to our patients. Firstly, we should consider whether the study results are valid. For articles about the effectiveness of healthcare interventions, the 3 key validity issues are whether the patients were randomly assigned to different treatments, whether they were analysed according to the groups to which they were assigned, and the extent of follow up. Secondly, if we judge the study to be valid, we examine the study results to determine whether the new treatment is effective, the size of the effect, and whether the effect is clinically important. When determining the clinical significance of effective treatments, findings can be expressed in 3 ways: as a change in relative risk, change in absolute risk and number needed to treat (NNT). Abstracts in Evidence-Based Nursing that describe effective treatments include these numbers, when data permit their calculation. The third step, the application to an individual patient, requires knowledge about both the study and the patient. This involves consideration of both the extent to which the patient resembles those who were enrolled in the study and the patient’s risk for the event for which the treatment was designed. This notebook will explain the concepts that help us to determine whether study findings should be applied to our own individual patients.

Let’s work through a randomised controlled trial abstracted in this issue of the journal (p52) that evaluates the effectiveness of a cognitive behavioural family intervention in reducing psychological distress and depression in caregivers of patients with Alzheimer’s disease. Addressing first the validity of this trial, we find that patient-caregiver dyads were randomly assigned to the 14 session family intervention or to 1 of 2 control groups (1 of which received a cathartic interview and 1 of which did not receive the family intervention or the interview), the patients were analysed in the groups to which they were assigned, and loss to follow up at 3 months was only 2%.

Turning to the results, let’s compare the intervention group with the no interview control group. We learn that within 3 months, 23% of the caregivers who received the family intervention met criteria for psychological distress (we will call this the experimental event rate or EER), whereas 77% of those in the no interview control group met the same criteria (we will call this the control event rate or CER). This difference was statistically significant.

Now that we know the results, let’s examine the application to our patients. We begin with the traditional measure of effect, which is the relative risk reduction (RRR), defined as the proportional reduction in rates of bad outcomes between experimental and control participants in a trial, calculated as (CER–EER)/CER. The RRR is (77%–23%)/77% or 70%; that is, the family intervention reduced the risk of caregivers experiencing psychological distress by 70%.

Why not confine our description of the clinical significance of this result to the RRR? The reason is that the RRR fails to discriminate huge absolute treatment effects from those that are trivial. For example, if the rates of psychological distress were 10 times less than those observed in this trial, and only 7.7% of control group caregivers and 2.3% of treatment group caregivers experienced psychological distress, the RRR would still be 70%. This is because the RRR ignores how rarely or commonly the outcome in question occurs anyway in the patients entering the trial (known as baseline risk); as a result, these measures cannot discriminate huge benefits and risks from small ones. The effect of treating low risk caregivers will differ from the effect of treating caregivers at a much higher risk of psychological distress.

In contrast to the non-discriminating RRR, the absolute differences in the rates of psychological distress between control and experimental group caregivers (ie, CER–EER) clearly do discriminate between these extremes; this measure is called the absolute risk reduction (ARR) and is defined as the absolute arithmetic difference in rates of bad outcomes between experimental and control participants in a trial. In the caregiver trial, the ARR = CER–EER = 77%–23% = 54%. In the hypothetical example provided above, in which 2.3% of experimental group caregivers and 7.7% of control group caregivers had psychological distress, the ARR = CER–EER = 7.7%–2.3% = 5.4%. The absolute differences take into account the baseline risk of patients and provide more detailed information than RRRs. But unlike the RRRs that can be recalled as whole numbers, ARRs are decimals and therefore more difficult to remember.

If however, we divide the ARR into 1 (ie, if we invert the ARR or take its reciprocal), so that it becomes 1/ARR, we generate a very useful number that represents the number of patients we need to treat (NNT) with the experimental treatment for the duration of the trial to prevent 1 additional bad outcome (eg, psychological distress). In the caregiver trial, we generate the number of caregivers we would need to treat with the family intervention to prevent 1 additional caregiver from experiencing psychological distress within 3 months of the intervention. The NNT is 1/ARR or 1/54% or 1/0.54 = 1.85; we usually round this number upward (because we can’t have part of a person), and we can now say that for every 2 caregivers who receive the family intervention, 1 case of psychological distress will be averted. Using the hypothetical example provided above in which the ARR was 5.4% rather than 54%, the NNT is 1/ARR or 1/0.054 = 18.5 or 19 caregivers who would need to receive the family intervention to prevent 1 additional caregiver from experiencing psychological distress. The NNT is higher for patients at lower risk of the health problem.

How can NNTs help me in clinical decision making? The NNT is a useful measure of the clinical effort that clinicians and patients must expend to help them avoid bad outcomes (eg, psychological distress) or experience good outcomes (eg, healing of a pressure sore). It is a meaningful way of expressing the magnitude of a treatment effect over a control. Knowing the NNT helps clinicians to determine whether the likely treatment benefits are worth the potential harm and costs. For example, we would be comfortable treating 10 patients with a safe, low cost treatment to prevent one patient from experiencing a pressure ulcer (an NNT of 10), but more reluctant to treat 10 000 patients with a risky, high cost treatment to prevent one patient from experiencing a pressure ulcer (an NNT of 10 000).
The rest of this editorial will focus on important points to keep in mind when calculating or interpreting NNTs.

**NNTs are only useful for interventions that produce dichotomous outcomes**

Because we are calculating the number of people who need to be treated, the only outcomes that lend themselves to this are those that count the number of people who experience the outcome (e.g., alive or dead, recovered or not recovered, healed or not healed). NNTs cannot be calculated when the outcome is presented as a mean value such as mean blood pressure or mean length of stay. In the caregiver trial described above, the outcome was the number of caregivers who experienced psychological distress, an outcome that lends itself nicely to the calculation of NNT.

**NNTs should always be interpreted in the context of their precision**

Because NNTs are only estimates of truth, NNTs presented in our abstracts are always accompanied by 95% confidence intervals or the limits within which the true NNT lies 95% of the time. Because estimates of precision are directly dependent on sample size, the smaller the number of patients in the study, the wider the confidence interval around the NNT. In the caregiver trial, the 95% confidence limits ranged from 2 to 7, indicating that the true NNT may be as low as 2 and as high as 7. Given that the true NNT could fall anywhere within the confidence interval, the decision to implement the treatment should be based on consideration of the outer limits of the confidence interval.

**Interpretation of NNTs must always consider the follow up time associated with them**

Because the number of reported events in a study has occurred by following up the study patients for a specified period of time, this must be reflected in the interpretation of the NNT. For example, in the caregiver study, patients were followed up for 3 months. The NNT for psychological distress at 3 months was 2 (95% CI 2 to 7). Put into words, 2 caregivers would need to receive the family intervention to prevent 1 additional case of psychological distress at 3 months and the true NNT could be as low as 2 and as high as 7.

**Clinical decision making must consider adverse outcomes as well as positive effects**

Treatments with positive effects may often have adverse effects as well. To determine the effect of the adverse events, we calculate the number needed to harm or NNH, which is defined as the number of patients who, if they received the experimental treatment, would lead to 1 additional person being harmed compared with patients who receive the control treatment. Like the NNT, NNH is calculated as 1/absolute difference and is accompanied by a confidence interval.

A study abstracted in the July 1999 issue of *Evidence-Based Nursing* that evaluated the effectiveness and safety of pressure bandages applied immediately after coronary angiography illustrates the need to consider both the benefits and adverse effects of a treatment. Within 6–12 hours after coronary angiography, only 3.5% of patients assigned to the pressure bandage group (EER) had bleeding compared with 6.7% of patients assigned to the no bandage control group (CER). This difference was statistically significant. The RRR is (CER–EER)/CER = (6.7%–3.5%)/6.7% = 48%, meaning that pressure bandages decreased the relative risk of bleeding after coronary angiography by 48%. The ARR is 6.7%–3.5% = 3.2%. The NNT is 1/0.032 = 32, which means we would need to treat 32 people with pressure bandages for 6–12 hours after coronary angiography to prevent 1 additional person from experiencing bleeding.

Although this appears to be an important positive effect, it must be considered in conjunction with the adverse effects. Patients in the bandage group had a higher incidence of nausea, back pain, groin pain, leg pain, and urinary difficulties. Looking more closely at groin pain, 17.5% of patients in the bandage group and 4.7% of patients in the control group experienced groin pain during the 6–12 hour time period. This absolute risk increase of 12.8% generates an NNH over 6–12 hours of 8, meaning that we only need to treat 8 patients with pressure bandages for 6–12 hours to cause 1 additional patient to have groin pain.

Clinicians and patients must decide when treatment effects are large enough to more than offset the adverse effects of a treatment. Research has been done to estimate the threshold NNT or the point at which the therapeutic risk equals the therapeutic benefit, however such discussion is beyond the scope of this notebook.

**NNTs will vary with baseline risk**

Because NNTs vary with baseline risk, we need to estimate the baseline risk of our own untreated patients relative to the average control patient in the trial. Let’s consider 2 hypothetical examples to illustrate how the baseline risk of our own patients may influence our decision to implement an effective intervention. The first focuses on the prevention of adolescent pregnancy. Let’s say that a study has been completed in the UK that shows the effectiveness of an adolescent pregnancy prevention programme. Two nurses, 1 in the US and 1 in the Netherlands, are considering whether to implement this programme in their countries. The NNTs will vary dramatically in these 2 countries because the baseline risk of adolescent pregnancy in the US is the highest of all developed countries, whereas the baseline risk of adolescent pregnancy in the Netherlands is one of the lowest in the world. As a result, the NNT to prevent 1 additional pregnancy in the US will be dramatically lower than the NNT to prevent 1 additional pregnancy in the Netherlands. Consequently, the nurse in the US might justifiably decide to go ahead with the intervention, whereas the nurse in the Netherlands might be equally justified in choosing not to implement the programme.

The second example considers neighbourhoods. Consider a study that has shown the effectiveness of a home visiting intervention to improve parenting skills and prevent child abuse. There might be certain neighbourhoods in a city where the incidence of child abuse is much higher than in other neighbourhoods. By considering baseline risk and calculating NNTs for specific neighbourhoods, decisions can be made about whether the entire city can and should receive the intervention or whether only high risk neighbourhoods should receive the intervention.

Our patient’s expected event rate can be estimated in various ways. Firstly, we can assign our patient the same event rate as that experienced by the control group in the trial. Although this is simple, it is only sensible if our patient is very similar to the average control group patient. Secondly, if the study presents the data for subgroups of patients, and one subgroup shares similar characteristics with our patients, we can assign our patient the control group rate for that subgroup. Thirdly, we can look for a prospective study that examined the prognosis of untreated patients like ours and use its results to assign a baseline risk rate to our patient.  

For patients at very high risk of the target event, the NNT will tend to be low, and treatment is likely to be justified. For patients at very low risk of the target event, the NNT is likely to be high enough to raise doubts about whether treatment is warranted, even when the outcome being prevented is serious. Chatellier et al describe a systematic review about coronary artery bypass graft surgery in patients with stable coronary heart disease that illustrates the importance of establishing our patient’s baseline risk when determining the NNT. The 5 year mortality was 6.3%, 13.9%, and 25.2% in patients with the lowest, middle, and highest risk, respectively. Assuming that the same 39% reduction of the 5 year risk of mortality existed in each subgroup, the NNT to avoid 1 death in the lowest, middle, and highest risk groups would be 40, 18, and 10, respectively.

Once we have estimated our patient’s baseline risk, the NNT can be calculated in 2 ways. The first makes use of a nomogram designed by Chatellier et al. To use this nomogram, a straight line is drawn from the point corresponding to the proportion of events in our patient on the left hand scale (absolute risk in the absence of treatment) to the point corresponding to the relative risk reduction calculated from the trial on the central scale. The point of intercept of this line with the right hand scale gives the NNT. By taking the upper and lower limits of the confidence interval of the RRR we can then obtain the upper and lower limits of the NNT. This allows us to assess the precision of the result and the magnitude of effectiveness on the most optimistic and most pessimistic limits of the confidence interval.

Because we may not always have the nomogram with us, an alternate method to calculate NNT might be preferable. We can determine the relation between our patient’s baseline risk and that of the average control patient in the trial. The relation is expressed numerically as a decimal fraction we will call “F.” For example, if our patient has twice the baseline risk as the average control patient, then F = 2; if our patient has half the baseline risk, then F = 0.5; if our patient has the same baseline risk, then F = 1. The NNT for our patient is simply the reported NNT divided by F. Going back to our caregiver hypothetical example in which rates of psychological distress were 10 times less than the baseline risk of control group caregivers in the trial, then F = 0.1 and NNT/F = 2/0.1; thus, 20 caregivers with a lower baseline risk would need to receive the family intervention to prevent 1 additional caregiver from experiencing psychological distress within 3 months.

**Nomogram for calculating the number needed to treat.** Reprinted with permission from BMJ 1996;312:426–9.

**Costs must be considered in clinical decision making**

Certain effective treatments may not be viewed as sufficiently cost effective to warrant implementation.

**Conclusion**

When faced with a study that evaluates a nursing intervention, we now know how to determine if the study results are valid. If valid, we know how to find and interpret the study findings. The final step is determining whether the findings of the study should be applied to our patients. It is at this point that the NNT helps us to translate statistical significance into clinical significance.

ALBA DICENSO, RN, PhD
School of Nursing and
Department of Clinical Epidemiology and Biostatistics
Faculty of Health Sciences, McMaster University
Hamilton, Ontario, Canada

**Patient preferences must be elicited in clinical decision making**

Once the patient has been informed about the benefits and risks of a specific treatment, the patient’s preferences or values should guide the treatment decision.
In this article, we use RRR to illustrate treatment effects. 2 other terms which are used to illustrate treatment effects are: *relative benefit increase* (RBI), defined as the proportional increase in rates of good outcomes between experimental and control participants in a trial. Both of these are calculated identically to RRR (ie, CER-EER/CER).

Although we use ARR to illustrate treatment effects in this paper, 2 other terms which reflect absolute differences are *absolute benefit increase* (ABI), defined as the absolute arithmetic difference in rates of good outcomes between experimental and control participants in a trial, and *absolute risk increase* (ARI), defined as the absolute arithmetic difference in rates of bad outcomes between experimental and control participants in a trial. Both of these are calculated identically to ARR (ie, CER-EER).