Applying the results of trials and systematic reviews to our individual patients

To translate evidence into clinical practice, clinicians need to judge how to apply the evidence to individual patients. For a complete discussion on how to assess the validity and applicability of therapeutic studies, we refer you to the *Users' guides series.*

In this EBMH Note, we will discuss 4 questions that clinicians might find useful when considering applicability, and we will highlight our discussion with the following clinical scenario:

We see a 77 year old man in clinic who was recently diagnosed with Alzheimer's disease. His medical history is significant for a seizure disorder that is well controlled with dilantin. On examination, he scores 22/30 on the Mini-Mental State Examination. His wife found some information about donepezil while surfing the internet and wants to know if her husband could receive this treatment. Together, we decide to review the evidence and judge whether he is a candidate for donepezil to slow the deterioration in his cognitive status. Using the terms “donepezil” and “dementia,” we find a relevant review in the Cochrane Library.

Is my patient so different from those in the study that the results cannot be applied?

To rigidly apply the inclusion/exclusion criteria of a trial when extrapolating its results may result in harm. It is generally more appropriate to consider whether the underlying pathobiology in our patient is so different that the study cannot give any guidance about management. For most differences in patient groups, the answer to this question is “no,” and we should instead think about how these differences might shift the balance between the benefits and harms of treatment. Differences between our patients and those in the trials tend to be qualitative (eg, matters of degree in risk or responsiveness) rather than quantitative (no response or adverse response to treatment). This rule has a few exceptions because the same disease may affect patients differently in important pathobiological ways. For example, a systematic review found that tricyclic antidepressants have little effect in children or adolescents.

For our patient, we need to consider whether any factors might cause donepezil to exert a qualitatively different effect in him. We know that all of the studies in the systematic review included patients with mild to moderately severe dementia, and our patient's disease can be considered mild. If our patient had severe dementia, however, the effect size from donepezil might be different. Similarly, we need to consider whether any factors could increase our patient's risk of an adverse event from donepezil. He is known to have a seizure disorder, and it has been noted that donepezil may increase the risk of seizures.

Is the treatment feasible in my setting?

Even if good evidence exists to support a management strategy, we need to consider whether we can provide it to our patient. Barriers to provision include geographic, economic, and organisational constraints. For example, even if a medication is available, we can only provide our patients with what they or the state can afford.

What are the likely benefits and harms from the treatment for my patient?

Once we have decided that the results of a study are applicable and feasible we need to individualise the benefits and risks to our patient. One method for summarising the results of randomised trials is the number of patients that we would need to treat with this therapy to prevent 1 additional adverse outcome (known as the number needed to treat [NNT]). The NNT is calculated as the inverse of the absolute risk reduction (ARR) and can also be calculated from an odds ratio (OR). When the OR is < 1 we can use the equation: \[ \frac{1}{\text{OR} - 1} \] where CER is the control event rate. From the systematic review that we found, an improvement in global clinical state shown on the Clinician's Interview-Based Impression of Change (CIBIC plus) score was observed with donepezil, 10mg/day (OR 0.58, 95% CI 0.26 to 0.56). Translating this, we would need to treat 6 people with donepezil for 24 weeks to prevent 1 person from having an unchanged or worsening CIBIC plus score.

Analogous to the NNT, the number needed to harm (NNH) is an expression for the number of patients who need to receive the intervention to cause 1 additional adverse event. It can be calculated as the inverse of the absolute risk increase (ARI) or from the OR using: \[ \frac{1}{\text{OR} - 1} \] where CER is the control event rate. From the systematic review that we found, an improvement in global clinical state shown on the Clinician's Interview-Based Impression of Change (CIBIC plus) score was observed with donepezil, 10mg/day (OR 0.58, 95% CI 0.26 to 0.56). Translating this, we would need to treat 6 people with donepezil for 24 weeks to prevent 1 person from having an unchanged or worsening CIBIC plus score.

The average NNT/NNH may not be directly applicable to an individual patient, and we need to consider what our patient's baseline risk of an event is to determine a patient specific NNT/NNH. If a study reports the risk for various subgroups, we can use the baseline risk for the subgroup most like our own patient. Single studies are often not large enough to provide us with this information, and we need to turn to systematic reviews for these data. Alternatively, we could derive an estimate of baseline risk from a clinical prediction guide or from articles that describe the prognosis of similar, untreated patients. Once we have identified the baseline risk, we can use it to calculate the NNT as \[ \frac{1}{\text{PEER} \times \text{RRR}} \] where the PEER is the patient expected event rate and the RRR is the relative risk reduction. In these calculations we are assuming that the RRR is constant across the range of baseline risks we encounter in clinical practice. If information is available from studies suggesting that the RRR may be different for various subgroups, then this RRR should be used in calculating the NNT.

A second approach for generating patient specific estimates is to use clinical judgment. We estimate the patient's risk of the outcome event relative to that of the average control patient in the study and convert this to a decimal fraction (f). Patients judged to be at less risk than those in the trials will be assigned f < 1 and those thought to be at greater risk will be assigned f > 1. Preliminary data suggest that experienced clinicians are accurate in estimating relative differences in baseline risk.

To calculate the patient specific NNT, we need to divide the average NNT by \( f \). For example, if we felt that our patient was at one half \( (f = 0.5) \) the risk of global clinical deterioration compared with the average patient in the study we identified, we would calculate his patient specific NNT as \( \frac{6}{0.5} = 12 \). No impor-
tant subgroups were identified in the systematic review that we found.

Similarly, we can use this method to generate a patient specific NNH using the term $f_l$. For example, we might think that our patient has twice ($f_l = 2$) the risk of an adverse event as the patients in the systematic review and his NNH becomes $6 (12/2)$.

**How will my patient’s values influence the decision?**

Regardless of how involved with the decision making the patient wants to be, it is essential for the clinician to explore the patient's values about the treatment and its potential risks and benefits. Our patient's values can be elicited in informal ways during exploratory discussions or by more formal (and time consuming) methods such as the time tradeoff, standard gamble, or rating scale techniques.6

There are several approaches to shared decision making support. For example, a clinical decision analysis (CDA), incorporating the patient's likelihood of the outcome events with his own values for each health state could be used. Doing a CDA for each patient, however, would be too time consuming for the busy clinician, and this approach therefore relies on finding an existing CDA. Because substantial variation often exists in values among individuals, CDAs that rely on group averages for values may not always be applicable to a particular patient,15–17 although the utility sensitivity analyses in a CDA may provide some guidance.

Decision aids that provide information about the target disorder, the management options, and the potential outcomes are becoming increasingly available.18 Efforts have also been made to develop alternative methods of presenting information to patients while incorporating patient values.15–18 Although all of these methods have merit, they sometimes fall short in comprehensibility, applicability, and ease of use in busy clinical settings.

Although not fully tested yet, 1 method of incorporating patients’ values in the decision making process is the likelihood of being helped or harmed (LHH).19 To calculate the LHH for our patient, $1/NNT (ARR)$ and $1/NNH (ARI)$ are combined into an aggregate ratio. (Alternatively we could use the ARR and the ARI in these calculations; in a pilot study, however, we found that physicians made fewer errors with the NNT/NNH than with the ARR/ARI in calculations.) For our patient the first approximation of the LHH is $(1/NNT) : (1/NNH) = (1/6) : (1/11) = 2$ to 1 in favour of donepezil.

We can particularise the LHH for our patient using the “f” factors described above. If we thought that his risk of an adverse event with donepezil was twice that of the patients in the trials, the risk adjusted LHH would be: $[(1/NNT) \times f_l] : [(1/NNH) \times f_l] = (1/6) : (1/11) \times 2 = 1$ to 1. We can tell the patient that the treatment is just as likely to help him as to harm him.

We need to explore our patient’s values to incorporate them into the LHH. Patients are provided with descriptions of the target event we are hoping to prevent and of the potential adverse event from the treatment. The clinician presents the patient with a rating scale (anchored at 0 [death] and 1 [full health]) and asks him to mark the value of the target event. Our patient and his wife assign a value of 0.1 to deterioration in global clinical status and a value of 0.95 to an adverse event from the medication. Using the 2 ratings we can infer that our patient believes deterioration to be 18 times worse than an adverse event from the medication ($1-0.1)/(1-0.95)$. We call this number the severity factor or $x$. Incorporating our patient’s values in the LHH, it becomes: $[(1/NNT) \times f_l \times x] : [(1/NNH) \times f_l] = [(1/6) \times 18] : (1/11) \times 2 = 18$ to 1 in favour of donepezil treatment.

Currently, this formulation is inexact, and we do not know how much difference it makes to patients or their clinical outcomes. We have presented a simple formulation for the LHH, but it could be modified to include other outcomes from treatment and indeed it could be used to compare various treatments.

Clearly, strides have been made in helping clinicians to apply evidence to individual patients. More research is needed, however, to develop and test intelligible, efficient methods of shared decision making that can be used at the bedside or in the busy clinic setting.

SHARON E STRAUS, MD
University of Toronto
Toronto, Ontario, Canada

FINLAY M’CALISTER, MD
University of Alberta
Edmonton, Alberta, Canada


2 Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users’ guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? JAMA 1993;270:598–601.

3 Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users’ guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me caring for my patients? JAMA 1994;271:59–63.


19 Straus SE. Individualizing treatment decisions: the likelihood of being helped or harmed. Evaluations and Health Professions. In press.