Part V: Using systematic reviews in practice
19 Applying the results of systematic reviews at the bedside

FINLAY A McALISTER

Summary points

Applying results from systematic reviews to an individual patient involves consideration of:

- **The applicability of the evidence to an individual patient**
  
  While some variation in treatment response between a patient and the patients in a systematic review is to be expected, the differences tend to be quantitative rather than qualitative.
  
  Outcomes research generally confirms that therapies found to be beneficial in a narrow range of patients have broader application in actual practice.

- **The feasibility of the intervention in a particular setting**
  
  Involves consideration of whether the intervention is available and affordable in that setting and whether the necessary expertise and resources are locally available.

- **The benefit: risk ratio in an individual patient**
  
  As a first step, the overall results must be summarised. In order to facilitate their extrapolation to a specific patient, formats which incorporate baseline risk and therapeutic effects (such as number needed to treat or number needed to harm) are preferable.
  
  Secondly, the results can be extrapolated to a specific patient by either considering results in the most relevant subgroup, or by multivariate risk prediction equations, or by using clinical judgement to determine a specific patient’s risk status.

- **The incorporation of patient values and preferences**
  
  This is an evolving field and current techniques include patient decision support technology or the expression of likelihood to help or harm.
Systematic reviews provide the best estimates of the true effects (both beneficial and adverse) of medical interventions, and other chapters in this book have outlined criteria for their performance and critical appraisal. However, these effect estimates are derived by collating data from a diverse range of patients. In this chapter, we build upon previous work and describe a framework for the clinician faced with deciding whether and how the results of a systematic review are applicable to a particular patient (see Summary points).

**Determining the applicability of the evidence to an individual patient**

Rather than pouring over the inclusion and exclusion criteria of the included studies to determine whether a particular patient would have been eligible, this question is better approached by asking “is the underlying pathobiology in my patient so different that the study cannot give any guidance?”. This involves consideration of the pathogenesis of the disease process as well as patient-specific biology and environmental exposures. While the potential challenges to the applicability of a systematic review described below may seem daunting, it is helpful to keep in mind that “differences between our patients and those we read about in trials tend to be quantitative (e.g. matters of degree in risk and responsiveness) rather than qualitative (no response or adverse response)”.

For example, although randomised clinical trials had clearly established that beta-blockers were beneficial in reducing mortality risk in acute myocardial infarction (MI), these drugs have been systematically under-used in MI patients as clinicians have tended to restrict their use to “ideal patients” who would have fulfilled trial entry criteria. However, observational studies in the real-world setting have suggested that beta-blockers are just as beneficial in patients who would have been excluded from the original trials because of perceived contra-indications (such as peripheral vascular disease, diabetes mellitus, congestive heart failure, or chronic obstructive airways disease). Moreover, such outcomes research has confirmed that the relative survival benefits are in the order of 30–40% in all major patient subgroups, including those traditionally under-represented in trials (such as women, blacks, or the elderly).

However, there are some exceptions to this general rule. First, while disease pathophysiology is usually similar in two or more patients with the same diagnosis, this may not always be the case. To the extent that differences exist, a systematic review’s applicability may be limited. For example, although the vast majority of the patients enrolled in the trials of angiotensin-converting enzyme (ACE) inhibitors in heart failure had systolic dysfunction, population-based cohort studies suggest that...
30–40% of all patients with heart failure have diastolic dysfunction. Thus, while a systematic review of these trials\textsuperscript{10} demonstrated a marked survival benefit from ACE inhibitor therapy, it is uncertain whether the benefits extend to heart failure patients with diastolic dysfunction (indeed outcomes research\textsuperscript{11} has failed to find any mortality advantage with ACE inhibitors in diastolic dysfunction). While the resolution of this clinical dilemma awaits further research, this example serves to highlight the necessity for a sound understanding of disease pathophysiology in order to properly interpret research evidence. Second, several patient-related factors can also impact upon the applicability of research results. For example, differences in drug metabolism (such as acetylation rates) or immune responsiveness among patients, arising from genetic polymorphism, may modulate the effects of interventions.\textsuperscript{12,13} Third, environmental factors may influence treatment effects (for instance, the frequency of thyroid dysfunction with amiodarone varies with environmental iodine intake).\textsuperscript{14} Finally, the balance between benefit and harm from an intervention may differ if an individual patient is less (or more) compliant than those in the studies: for example, non-compliant patients are at higher risk of bleeding with chronic warfarin therapy than their compliant peers (relative risk 2.3, \(P = 0.003\)).\textsuperscript{15}

Given these myriad influences on the applicability of research evidence, one should expect some variation in treatment response between patients in one’s practice and those described in systematic reviews. However, these variations are not always important (for example, the management of cataracts is generally similar despite the varied pathogenesis) or non-remediable (the dose of a drug can be adjusted based on individual patient responsiveness).\textsuperscript{3} To return to the example of bleeding with chronic warfarin therapy, cohort studies have shown that the complication rates seen in randomised trials can be achieved with compliant patients and clinicians in actual practice.\textsuperscript{16–18} Thus, the assumption that study results are applicable to a broader range of patients than enrolled in the trials generally holds true and outcomes research in the cardiovascular field consistently demonstrates that less harm results from this assumption than from withholding efficacious therapies from subgroups of patients not investigated in randomised trials.\textsuperscript{19}

**Determining the feasibility of the intervention in a particular setting**

In deciding whether the results of a systematic review can be extrapolated to an individual patient, the realities of local circumstance must also be considered. For example, is the intervention available and affordable in that setting? Thus, while there is little doubt that thrombolytic therapy confers significant benefits in patients with acute myocardial infarction,\textsuperscript{20} the cost
makes provision of this therapy prohibitive in many developing countries. Secondly, one must consider whether the necessary monitoring facilities are available if the intervention is offered. For example, if there is no facility for monitoring prothrombin times, it would be unwise to prescribe warfarin for a patient with atrial fibrillation, despite the trial evidence proving substantial efficacy. Finally, one must consider whether local expertise is sufficient to warrant provision of the intervention. For example, although a recent systematic review of carotid endarterectomy in patients with asymptomatic carotid stenosis concluded that surgery “unequivocally reduces the incidence of ipsilateral stroke”, the rate of perioperative complications in the trials was much lower than seen in audits of non-trial surgeons. As a result, carotid endarterectomy in centres with higher surgical complication rates will confer more harm than benefit. Thus, the clinician may decide to refer eligible patients to a centre with lower perioperative morbidity rates.

**Determining the benefit : risk ratio in an individual patient**

If the results of a systematic review are judged applicable and feasible, the clinician must then evaluate the likely benefits and harms from the intervention. This involves two steps: deriving clinically useful estimates of the overall results and extrapolating from the overall results to derive estimates for the individual patient. To illustrate this process, an example is outlined in Box 19.1 (adapted from Glasziou et al.) and will be referred to below.

**Deriving clinically useful estimates from the overall results**

Although the results of randomised clinical trials and systematic reviews with binary outcomes can be expressed in a number of ways, expressing the effects of treatment in terms of the number of patients one would need to treat to prevent one clinical event (NNT) is gaining widespread acceptance as the most relevant format for extrapolating to patients, can be directly applied to patients who are at the average risk in the included trials, and quickly adjusted at the bedside for patients who are not. The advantage to front line clinicians of the NNT over the more traditional reporting formats (such as relative risk reduction (RRR) or odds ratio (OR), which do not reflect baseline risk) is illustrated by considering the benefits of antihypertensive therapy for patients with varying degrees of blood pressure elevation (Table 19.1). As can be seen, the OR (and RRR) in all three blood pressure strata are approximately 40%, but the amount of effort required by clinicians and patients to prevent one stroke varies according to the baseline risk.

While NNTs are easily calculated when relative and absolute risks are reported (the NNT is the inverse of the difference in absolute event rates
Box 19.1 Determining the benefit:risk ratio in an individual patient – the example of warfarin therapy

Hypothetical patient
A 76-year-old female with hypertension and asymptomatic nonvalvular atrial fibrillation for at least three months. Transthoracic echocardiogram showed an enlarged left atrium, suggesting that attempts at cardioversion would be unlikely to be successful.

Deriving clinically useful estimates of the overall results
Benefit (defined as prevention of embolic stroke)
- control event rate (CER) 4·5% per year
- experimental event rate (EER) 1·4% per year
- absolute risk reduction (ARR) 3·1% per year
- NNT to prevent one embolic stroke 33 per year

Harm (defined as major bleeding)
- control event rate (CER) 1·0% per year
- experimental event rate (EER) 1·4% per year
- absolute risk increase (ARI) 0·4% per year
- NNH (i.e. to cause one major bleed) 250 per year

Extrapolating to the individual patient
1 Subgroup analysis:
   Benefit
   - CER in relevant subgroup 8·1% per year
   - EER in relevant subgroup 1·2% per year
   - ARR in relevant subgroup 6·9% per year
   - NNT in relevant subgroup 15 per year

   Harm
   - Event rates not reported in patient-specific subgroups.

2 Use of the $f$ factor
   Benefit
   - patient estimated to be at twice the risk of embolic stroke of the average patient in the trials; thus, $f$ factor = 2
   - average NNT/$f$ factor = patient-specific NNT; thus, patient-specific NNT = 33/2 = 17

   Harm
   - patient estimated to be at twice the risk of bleeding as the average patient in the trials; thus, $f$ factor = 2
   - average NNH/$f$ factor = patient-specific NNH; thus, patient-specific NNH = 250/2 = 125

NNT = number needed to treat; NNH = number needed to harm.
Table 19.1 Methods of reporting the results of systematic reviews.

<table>
<thead>
<tr>
<th>Patient strata (by diastolic blood pressure [DBP])</th>
<th>Stroke rates</th>
<th>Odds ratio</th>
<th>Relative risk reduction $[\text{RRR} = (P_c - P_A) / P_c]$</th>
<th>Absolute risk reduction $[\text{ARR} = P_c - P_A]$</th>
<th>Number needed to treat for five years to prevent one stroke $[(1/\text{ARR})]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBP &lt; 110 mm Hg</td>
<td>0.0148</td>
<td>0.0087</td>
<td>0.41</td>
<td>0.006</td>
<td>164</td>
</tr>
<tr>
<td>DBP ≥ 115 mm Hg</td>
<td>0.0201</td>
<td>0.0122</td>
<td>0.42</td>
<td>0.008</td>
<td>125</td>
</tr>
<tr>
<td>DBP &gt; 115 mm Hg</td>
<td>0.0263</td>
<td>0.0156</td>
<td>0.41</td>
<td>0.011</td>
<td>91</td>
</tr>
</tbody>
</table>

Data from Collins et al.\textsuperscript{27}
between the control and experimental arms, see Chapter 2), they cannot be easily calculated from ORs. Since the value of the OR does not always reflect the RR (particularly when disease incidence is above 10%, see also Chapter 2), the clinician must employ a formula or consult Table 19.2 to derive the overall NNT from systematic reviews which report only the OR. The NNT for the prevention of one embolic stroke with warfarin therapy in the average patient with non-valvular atrial fibrillation (as determined in a systematic review of the atrial fibrillation trials) is approximately 33 (Box 19.1).

Analogous to the NNT, the number needed to harm (NNH) is an expression of the number of patients who would need to receive an intervention to cause one additional adverse event. The NNH is the inverse of the difference in absolute adverse event rates between the control and experimental arms. For example, the NNH (harm defined as a major bleed) for warfarin therapy was 250 in the systematic review of the atrial fibrillation trials (Box 19.1).

**Extrapolating to the individual patient**

While the NNT and NNH are clinically useful estimates of the average treatment effects in patients at the average risk in the included trials, they may not be directly relevant to an individual patient; thus, the clinician must extrapolate in one of three ways.

First, if the systematic review presents estimates of treatment effects in various subgroups, the clinician can extrapolate using the NNT and/or NNH from the subgroup most relevant to their patient. For example, a systematic review of antiplatelet agents for the prevention of non-fatal myocardial infarction revealed similar proportional treatment effects in trials of primary and secondary prevention (OR 29% and 35% respectively). However, the NNT varied markedly such that 200 patients without symptomatic cardiovascular disease would need treatment for five years to prevent one myocardial infarction while only 71 “high risk” patients (those with prior myocardial infarction, stroke, or other cardiovascular event) would require treatment for three years to have the same clinical impact. Returning to our running example, the systematic review of atrial fibrillation trials provided estimates of risk and treatment effects in various subgroups defined by baseline clinical features; the patient with non-valvular atrial fibrillation outlined in Box 19.1 has a baseline risk of embolic stroke of approximately 8% per annum without treatment, and warfarin therapy is associated with an 85% relative risk reduction. Thus, the NNT for this patient would be approximately 15 (Box 19.1). Unfortunately, the infrequent nature of adverse events precluded the investigators from determining the NNH in patient-specific subgroups. In this situation, one
Table 19.2 Deriving the NNT from the odds ratio.

<table>
<thead>
<tr>
<th>Control event rate</th>
<th>Preventive intervention</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
<td>0.55</td>
</tr>
<tr>
<td>0.05</td>
<td>41</td>
<td>46</td>
</tr>
<tr>
<td>0.1</td>
<td>21</td>
<td>24</td>
</tr>
<tr>
<td>0.2</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>0.3</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>0.4</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>0.5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>0.7</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>0.9</td>
<td>12</td>
<td>15</td>
</tr>
</tbody>
</table>

Reproduced from McQuay and Moore, with permission. The formula for determining the NNT for preventive interventions is
\[ \frac{1 - \left( \text{CER} \times (1 - \text{OR}) \right)}{(1 - \text{CER}) \times \text{CER} \times (1 - \text{OR})} \]. For treatment, the formula is \[ \frac{\text{CER} \times (\text{OR} - 1) + 1}{\text{CER} \times (\text{OR} - 1) \times (1 - \text{CER})} \].

CER = control event rate, OR = odds ratio.
is left with using the overall results or using one of the other methods of extrapolation outlined below.

While the use of subgroup analyses seems intuitively appealing, we must sound a note of caution at this point. A full discussion of the limitations of subgroup analysis is beyond the scope of this chapter, but has been covered in full elsewhere. In particular, one should be wary of systematic reviews which stratify patients by risk (as determined by post-hoc analysis of event rates in the control groups of the included trials) as these analyses may produce biased and inaccurate estimates of the relative treatment effects (see also Chapters 8 and 10). Instead, subgroups should be based on measurable patient characteristics at baseline (for example, gender, age, or primary versus secondary prevention). Since the derivation of such subgroups is often not possible using published trial results, this is a key advantage of individual patient data meta-analyses (see also Chapter 6).

Secondly, as an extension of the subgroup approach, one can use multivariate risk prediction equations to quantitate an individual patient’s potential for benefit (and harm) from therapy. For example, returning to the systematic review of endarterectomy for asymptomatic carotid stenosis referred to earlier, investigators are working on a prognostic model to identify patients who are most likely to benefit from operative intervention. This model incorporates the risk of stroke without surgery (and thus the potential benefit from surgery) with the risk of stroke or other adverse outcome from surgery. Application of this model to patients with symptomatic carotid stenosis enables clinicians to identify high-risk patients who benefit considerably from surgery (OR 0.12, 95% CI 0.05 to 0.29) from other patients with the same degree of stenosis but little to gain from surgery (OR 1.00, 95% CI 0.65 to 1.54). While these multivariate risk prediction models can be derived from the clinical trial data included in the systematic review, it is preferable if they come from other datasets such as population-based cohort studies. No multivariate risk prediction models have yet been published for chronic warfarin therapy in non-valvular atrial fibrillation.

Finally, in the absence of subgroup data or prognostic models, the clinician can employ clinical judgement by dividing the average NNT (or NNH) by a factor (f) which relates the risk of the individual patient to that of the average patient in the published reports. This factor is expressed as a decimal such that patients judged to be at less baseline risk than those in the trials will be assigned an \( f < 1 \) and those at greater risk will be assigned an \( f > 1 \). For example, if we did not have the subgroup-specific data for atrial fibrillation discussed above, we may have estimated that, since they are older than the average patient in the atrial fibrillation trials and have hypertension (while less than half of the randomised trial patients did), the patient outlined in Box 19.1 is at twice the risk of embolic stroke than the
average randomised trial patient. Thus, the patient-specific NNT would be about 17. By the same token, we may have estimated they were at twice the risk of major bleeding (because of their age). Thus, the patient-specific NNH would be 125.

While this method may appear overly subjective, recent empirical evidence suggests that clinicians are accurate in estimating relative differences in baseline risk (i.e. \( f \)) between patients (far exceeding our abilities to judge absolute baseline risks). However, it should be recognised that this method implicitly assumes that the proportional treatment effects (RR or OR) from an intervention are constant across different baseline risks, an assumption that may not hold for all therapies. Moreover, further research is needed into the basis for, and determinants of, clinical judgement.

Incorporating patient values and preferences

After deriving patient-centered estimates for the potential benefit and harm from an intervention, the clinician must integrate this with their patient’s values and preferences about therapy. Indeed, active patient involvement in medical decision making improves their quality of life and outcomes from treatment; moreover, there is preliminary evidence that it may also reduce health care expenditures. However, the optimal means of involving patients in treatment decisions has not yet been found and patient decision support technology is a rich vein for current research. Decision support technology is distinct from general patient education in its focus on the benefits and risks of alternatives (with explicit discussion of the probabilities and consequences of clinically important outcomes), the tailoring of the information to the particular patient’s risk profile, the emphasis on choice and shared decision making, and explicit elicitation of patient values.

Until the techniques of formal decision analysis have evolved to the stage that they are feasible to use at the bedside, interim techniques such as decision aids or the expression of likelihood to help or harm (a formula weighting the ratio of NNT:NNH by patient values) will serve this need.

Conclusion

While systematic reviews provide the best estimates of the true effects of an intervention, their application at the bedside is a “difficult, time-consuming, and incompletely studied skill”. In this chapter, we have outlined a framework for approaching this task and anticipate that ongoing research will greatly expand our understanding and performance of these steps.
Acknowledgements

FAM is supported by the Medical Research Council of Canada and the Alberta Heritage Foundation for Medical Research.

3 Dans AL, Dans LF, Guyatt GH, Richardson, S. Users’ guides to the medical literature. XIV. How to decide on the applicability of clinical trial results to your patient. JAMA 1998;279:545–9.
48 Straus SE, Sackett DL. The likelihood of being helped versus harmed – a new method of presenting information to patients about the risks and benefits of therapy. 3rd Cochrane Colloquium, Amsterdam, October 1997.
49 Ross JM. Commentary on applying the results of trials and systematic reviews to individual patients [EBM Note]. Evidence-Based Med 1998;3:167.