

## EBN users' guide

# Evaluation of systematic reviews of treatment or prevention interventions

In a previous article in this series we explained how the critical appraisal of research is an essential step in evidence-based health care because most published research is too poor in quality to be applied to clinical practice.<sup>1</sup> Critical appraisal is made easier through the use of quality checklists that can help you to appraise research studies systematically and efficiently. The 3 basic appraisal questions are the same whether the clinical question is about treatment, diagnosis, prognosis, or causation:

- Are the results of the study valid?
- What were the results?
- Will the results help me in caring for my patients?<sup>1-3</sup>

The first 2 articles in the EBN users' guide series focused on critical appraisal of primary studies of treatment or prevention.<sup>1,2</sup> This guide will deal with critical appraisal of systematic reviews, beginning with a clinical scenario and applying the appraisal questions to the review by Glazener and Evans on the effectiveness of alarm interventions for nocturnal enuresis in children.<sup>4</sup>

### Are the results of this systematic review valid?

- Is this a systematic review of randomised trials?
- Does the systematic review include a description of the strategies used to find all relevant trials?
- Does the systematic review include a description of how the validity of individual studies was assessed?
- Were the results consistent from study to study?
- Were individual patient data or aggregate data used in the analysis?

### What were the results?

- How large was the treatment effect?
- How precise is the estimate of treatment effect?

### Will the results help me in caring for my patients?

- Are my patients so different from those in the study that the results don't apply?
- Is the treatment feasible in our setting?
- Were all clinically important outcomes (harms as well as benefits considered)?
- What are my patient's values and preferences for both the outcome we are trying to prevent and the side effects that may arise?

*How to critically appraise review articles*

### Clinical scenario

You work in an outpatient paediatric clinic and often see children whose parents are frustrated and concerned about their child's night time bed wetting (nocturnal enuresis). They ask if you would recommend the use of night time alarm

systems that they have seen advertised. Should you recommend the use of these alarms for their children with non-organic nocturnal enuresis?

### The search

An efficient literature search strategy begins with a search for relevant research that has already met quality criteria and has been summarised. The evidence-based journals publish exactly this type of material, and you search the *Evidence-Based Nursing* web site ([www.evidencebasednursing.com](http://www.evidencebasednursing.com)) using the term "enuresis." You are in luck! You find an abstract and commentary for a systematic review by Glazener and Evans. The abstract, entitled "Review: alarm interventions reduce nocturnal enuresis in children," reports that alarms reduce enuresis compared with no intervention or other interventions (p110 of this issue). This looks promising and you decide to retrieve the full Cochrane review by searching the *Cochrane Library* using the text word "enuresis." You then need to make a judgment about the quality of the review—can you confidently use this review to inform practice?

### What is a systematic review?

Basing clinical decisions on a single research study can be a mistake for several reasons. Individual studies may have inadequate sample sizes to detect clinically important differences between treatments (leading to false negative results), the results of apparently similar studies may vary because of chance, and there are often subtle differences in the design of studies and the participants that may lead to different or even discrepant findings. A *systematic review* (often called an overview) is a rigorous summary of all the research evidence that relates to a specific question; the question may be one of causation, diagnosis, or prognosis but more frequently involves the effectiveness of an intervention. Systematic reviews differ from unsystematic reviews in that they attempt to overcome possible biases at all stages, by following a rigorous methodology of search, research retrieval, appraisal of the retrieved research for relevance and validity (quality), data extraction, data synthesis, and interpretation. One way in which bias is reduced is by the use of explicit, pre-set criteria to select studies for inclusion on the basis of relevance and validity. A second way is by having  $\geq 2$  people independently make study selection decisions, compare results, and discuss discrepancies before moving on to independently extract data from the studies. Explicit details of the methods used at every stage are recorded. Many, but not all, systematic reviews incorporate meta-analysis (the quantitative combination of the results of similar studies). Meta-analysis produces an overall summary statistic that represents the effect of the intervention across different studies. Because meta-analysis in effect combines the samples of each contributing study to create one larger study, the overall summary statistic is much more precise than the effect size in any one contributing study.

Systematic reviews have the potential to overcome several barriers to research utilisation in clinicians. Nurses have difficulties using research because of lack of access and time to retrieve

numerous research reports and lack of skills to appraise and synthesise the articles once retrieved. Systematic reviews offer nurses a solution in the form of a summary of research based knowledge on a topic, which takes into account the validity of the primary research. Nevertheless, not every systematic review is of high quality and the critical appraisal step remains essential.

### Are the results of this systematic review valid?

#### IS THIS A SYSTEMATIC REVIEW OF RANDOMISED TRIALS?

Good systematic reviews determine at the outset to include those studies that used the most appropriate design to answer the clinical question. Questions about the effectiveness of treatment or prevention are best answered by randomised controlled trials, whereas questions about harm or prognosis are best answered by cohort studies.<sup>5</sup> The review by Glazener and Evans includes 22 randomised trials: 5 compared alarms with no intervention and 17 compared different alarms, alarms with drugs or behavioural interventions, or combination interventions.<sup>4</sup>

#### DOES THE SYSTEMATIC REVIEW INCLUDE A DESCRIPTION OF THE STRATEGIES USED TO FIND ALL RELEVANT TRIALS?

It is usually necessary to search several electronic bibliographic databases and to use other strategies, such as journal handsearching and consultation with experts, to ensure that every primary study is identified. A search confined only to the Medline and CINAHL databases will be biased towards studies published in English, and those that found significant differences between interventions (if a reviewer only finds the 2 studies that found a difference and not the 10 that did not, she is likely to draw misleading conclusions).<sup>6</sup> Therefore, a search strategy is considered thorough if it includes several databases, if the reference lists of relevant papers were searched, if key journals were searched by hand, and if key informants were contacted. In the Glazener review, 11 different electronic databases were searched.<sup>4</sup> As another step to ensure that all possible studies were included and to minimise publication bias,<sup>6</sup> organisations, manufacturers, researchers, and health professionals concerned with enuresis were contacted to identify unpublished studies. References were also checked for additional studies.<sup>4</sup> Handsearching of key journals was not reported for this review. Handsearching contributes to the completeness of retrieval, as occasionally those who index articles for databases such as Medline may use inappropriate keywords or may miss articles or even whole journal issues. This particular review could be considered to have an extensive and thorough search strategy, with the exception of the lack of handsearching.

Every systematic review should grow from a focused question, which itself leads to the development of inclusion criteria and a sensitive and specific search strategy. Once the articles are retrieved, it is necessary to be pedantic about the *application* of inclusion and exclusion criteria. The review by Glazener and Evans identifies clear, focused questions: *do alarm interventions reduce nocturnal enuresis? Are alarm interventions more effective than other interventions?* Inclusion criteria were that children had to be randomised to alarm treatment or to a control condition, which could involve no intervention or other behavioural methods or drugs for nocturnal enuresis. Trials of children with organic causes for bed wetting were excluded, as were trials of daytime wetting.<sup>4</sup>

#### DOES THE SYSTEMATIC REVIEW INCLUDE A DESCRIPTION OF HOW THE VALIDITY OF THE INDIVIDUAL STUDIES WAS ASSESSED?

An unsystematic narrative review often reports on study findings without considering the methodological strength of the

studies. Differences in study quality might explain differences in results because studies of poorer quality tend to overestimate the effectiveness of interventions.<sup>7</sup> Quality rating scales are sometimes used in the analysis to make comparisons of outcome by study strength. If there are many studies to consider, the authors may decide to apply a quality rating threshold for inclusion of studies in the review, or to give greater weight to stronger studies.

A prespecified quality checklist helps to ensure that reviewers appraise each study consistently and thoroughly—another means of minimising bias. Having  $\geq 2$  raters, with some mechanism described for dealing with divergence of opinion, helps to reduce both mistakes and bias and increases the confidence of the reader in the systematic review. The quality rating tool usually includes criteria such as those presented in previous users' guides in this series.<sup>1 2</sup>

The review by Glazener and Evans applied explicit, prespecified validity criteria including the level of concealment of allocation at randomisation, the comparability of the groups at baseline, blinding of outcome assessment, use of intention to treat analysis, and the extent of follow up.<sup>4</sup> The reviewers were explicit about the quality criteria used in this review.

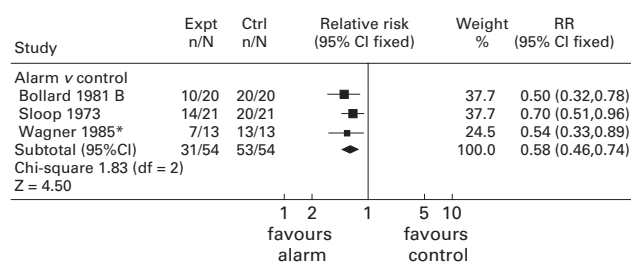
#### WERE THE RESULTS CONSISTENT FROM STUDY TO STUDY?

Although we would not expect to find the same magnitude of effect in all studies, we would be more confident in using the results of a review if the results of individual studies were qualitatively similar—that is, all showing a positive effect or all showing no effect. But what if treatment effect differs across studies? Many systematic reviews identify studies with important differences between them in terms of the types of patients included; the timing, duration and intensity of the intervention; or the outcome measures. The reviewers make decisions about whether meta-analysis is appropriate by using a combination of judgment and a statistical test for heterogeneity—a test of the extent to which differences between the results of individual studies are greater than you would expect if all studies were measuring the same underlying effect and the observed differences were only because of chance. The more significant the test of heterogeneity, the less likely that the observed differences are because of chance alone, indicating that some other factor (eg, study design, patients, intervention, or outcomes) is responsible for differences in treatment effects across studies.<sup>3</sup>

The value of judgment in deciding whether statistical synthesis is appropriate cannot be overemphasised, however, because the statistical test for heterogeneity has low power and may fail to identify important differences in study results. Readers of reviews must make judgments, using their clinical expertise, about whether meta-analysis makes clinical and methodological sense. In the review by Glazener and Evans, the authors chose to statistically combine 3 RCTs that compared alarms with no intervention for the outcome of “number not achieving 14 consecutive dry nights or relapsing” (this outcome will now be referred to as “treatment failure or relapse”).<sup>4</sup> All 3 trials had similar populations, interventions, and outcome measures, and the results were similar in direction and size of differences between the treatment and control groups. The test for heterogeneity was not significant.<sup>4</sup>

Different statistical approaches can be used in meta-analysis. *Fixed effects models* are most commonly used when no significant heterogeneity exists between studies; this model assumes that the true effect of the treatment is the same in each study, with differences in results arising because of chance. A *random effects model* assumes that the study results vary around some overall average treatment effect, and the calculation of the summary

Review: alarm interventions for nocturnal enuresis in children  
 Comparison: ALARM v CONTROL  
 Outcome: numbers not achieving 14 dry nights or relapsing



Graphical display of results of meta-analysis. Reprinted from Glazener CMA, Evans JHC. Alarm interventions for nocturnal enuresis in children. *Cochrane Database Syst Rev* 2001;(1):CD002911.

statistic incorporates an estimate of between study variation.<sup>8</sup> The Glazener review used a fixed effects model.<sup>4</sup>

#### WERE INDIVIDUAL PATIENT DATA OR AGGREGATE DATA USED IN THE ANALYSIS?

Less commonly, authors may request individual patient data from the investigators of individual studies. In this case, rather than using the study results (eg, relative risks or odds ratios), individual patient data may be combined across studies to allow comparisons of outcomes for subgroups, defined, for example, by age or severity of illness. The Glazener and Evans review did not include individual patient data in their results.<sup>4</sup>

#### What were the results?

##### HOW LARGE WAS THE TREATMENT EFFECT?

Comparing a simple count of the number of studies that found a positive effect of an intervention with the number that found no effect or a harmful effect, would be misleading. This would assume that all studies had equal validity, power, duration, dosage, etc, to have the potential for the same effect. Meta-analysis, when appropriate, can assign different weights to individual studies so that those with greater precision, or higher quality, will make a greater contribution to the summary estimate. Glazener and Evans used meta-analysis to determine the relative risk (RR) of treatment failure or relapse in the intervention versus the control group and used the graphical display common to Cochrane Reviews to summarise their findings (fig). It is useful to learn how to interpret these figures as they can portray much information at a quick glance.

Firstly, focus on the far left column. Looking down you will see a row for each of the 3 studies included in this comparison, referenced by the name of the first author and year the study was done. The list begins with the study by Bollard (1981 B) and ends with the study by Wagner (1985). Looking along each study (row), there is a box in the middle of the figure, with a horizontal line through it. It is possible to interpret whether an intervention is more or less effective than its comparator without reading the actual numbers. The box indicates the result of that particular study, in this case a RR, and the horizontal line indicates the 95% confidence interval (CI) around that RR. A RR of 1 would mean there is no difference in the event rate (ie, treatment failure or relapse) between the treatment and control groups; the risks are the same. A RR on the right side of the vertical line representing 1 (ie, a RR >1) would favour the control condition (ie, more children with treatment failures or relapses in the alarm group), and a RR on the left side of the line would favour the treatment (more children with treatment failures or relapses in the control group).

The reader can get an idea of the heterogeneity of the study results simply by looking at how the lines are scattered. In this

case, the results of each of the 3 studies are on the left side of 1—that is, all estimates of RR were < 1, and the CIs overlapped considerably. If the boxes for different studies were on both sides of 1 and had non-overlapping CIs, you would have less confidence in using the results of this review in a clinical decision. Applying this information, the study in the first row by Bollard (1981 B) has an associated RR on the left side of 1, indicating that fewer children in the alarm group had treatment failures or relapses than children in the control group. Because the CI does not cross the vertical line (ie, does not include 1), this difference is statistically significant. Looking down the list, all 3 studies reported a statistically significant reduction in treatment failures when alarms were used.

The result of combining all 3 studies is found at the bottom of the figure. The overall summary statistic, in this case the combined RR, is depicted as a diamond which also encompasses the 95% CI. The edges of the diamond do not cross or touch 1, indicating a statistically significant difference in favour of alarms. When outcomes are dichotomous (eg, alive or dead, dry or wet), meta-analyses generally use RRs or odds ratios for reporting the summary statistic; when the outcomes are continuous (eg, blood pressure, blood glucose, or weight), the calculation is the mean effect size or mean difference.<sup>9</sup> Each of these statistics may be weighted or unweighted. When mean effect size or mean difference are reported, the vertical line of no difference is at 0 rather than 1.

Although the summary statistic is the most important “bottom line,” more can be learnt from the figure. The second column from the left (Expt n/N) gives the number in the treatment group who experienced the outcome of interest (treatment failure or relapse) (n) out of the total number in the treatment (alarm) group (N). The third column (Ctrl n/N) gives the same information for the control group in each study. The fifth column of numbers (Weight %) tells you how much a particular study contributed to the overall summary statistic, with more weight given to studies of greater precision. The column on the far right [RR (95% CI fixed)] provides the RRs and accompanying 95% CIs for each individual study corresponding to the box and horizontal line for that study.

Rather than reporting RRs, *Evidence-Based Nursing* usually reports the relative risk reduction (RRR) and the number needed to treat (NNT), each with accompanying 95% CIs. Previous EBN notebooks explained these measures of effect in detail.<sup>8–10</sup>

From the table in the Glazener and Evans abstract, the RRR is 42%. In other words, the alarm intervention reduced the risk of treatment failure or relapse (a bad outcome) by 42% in the intervention compared with the control group. You can calculate this from the figure by subtracting the RR from 1 (ie,  $1 - 0.58 = 0.42$  or 42%). The 95% CI around the RRR of 42% is 26% to 54%, indicating that the true RRR may be only 26% or may be as large as 54%. You can calculate this from the figure by subtracting each end of the CI from 1 (ie,  $1 - 0.46 = 0.54$  or 54% and  $1 - 0.74 = 0.26$  or 26%). The NNT to prevent one additional treatment failure or relapse for alarms versus usual care was 3 (95% CI 2 to 4). In other words, to prevent 1 additional treatment failure or relapse (ie, not achieving 14 consecutive dry nights or relapsing after treatment completion), you would need to treat 3 children with alarms, and we are 95% certain that the true NNT may be as low as 2 and as high as 4 children.<sup>10</sup> One of the limitations of using NNTs derived from meta-analyses is that the patients entered into individual trials may vary considerably—particularly in terms of how susceptible they were to the outcome of interest.<sup>11</sup> In many reviews, the length of follow up varies across the primary studies, making the



NNT difficult to interpret. Furthermore, a decision about whether or not to use enuresis alarms would also involve consideration of risks, costs, and patient preference and acceptability.

#### HOW PRECISE IS THE ESTIMATE OF TREATMENT EFFECT?

As described in a previous users' guide,<sup>3</sup> CIs around the RR and the RRR indicate the precision around the estimate of the true treatment effect, which can never be really known. Wide CIs indicate less precision in the estimate. The convention is to use the 95% CI, which represents the range within which we are 95% certain that the true value lies.<sup>10</sup> Precision increases with larger sample sizes, although this is difficult to see in our example of the Glazener and Evans review because the 3 studies had similar sample sizes.

The CI for the final summary RR is fairly narrow (0.46 to 0.74).<sup>4</sup> The CI is useful for decision making because we can look at the limit closest to 1 (no effect) and ask ourselves "if the effect was as small as this, would it be worthwhile?" In this review, the risk, or probability, of treatment failure or relapse in the alarm group is 58% of the risk of treatment failure or relapse in the control group. The 95% CI indicates that the true risk of treatment failure or relapse in the alarm group may be only 46% of the risk in the control group, or the risk may be as high as 74%. To determine your confidence in adopting the intervention, you should consider the boundary of the CI closest to 1. If the risk of treatment failure really is approximately 75% of the risk in the control condition, given the cost, inconvenience, and side effects of the intervention, would you still want to implement the intervention?

#### Will the results help me in caring for my patients?

##### ARE MY PATIENTS SO DIFFERENT FROM THOSE IN THE STUDY THAT THE RESULTS DON'T APPLY?

As in the users' guide for evaluating a treatment,<sup>2</sup> you need to consider the characteristics of the patients within the individual studies included in a review, and how similar they are to your own. Are there reasons why the results would not be applicable to your patients? Data from the review by Glazener and Evans showed that participants were somewhat more likely to be boys and to have a mean age of about 9 years,<sup>4</sup> and so are similar to some of the patients you see in your clinic. The review also needs to provide enough information about the intervention to enable implementation.

##### IS THE TREATMENT FEASIBLE IN MY SETTING?

Use of nocturnal alarms seems a clinically feasible intervention in terms of care provider ability to recommend alarm systems. Feasibility for this scenario, however, would relate to the ability of parents to either buy the alarm system or have it provided to them.

##### WERE ALL CLINICALLY IMPORTANT OUTCOMES (HARMS AS WELL AS BENEFITS) CONSIDERED?

Researchers try to look at all outcomes, both positive and negative, that are a result of treatment and are important to the patient and the healthcare system. These might include mortality, morbidity, quality of life, cost effectiveness and patient satisfaction. Aspects of harm may not be systematically collected and fully reported in primary studies and few systematic reviews of treatments undertake thorough searches for relevant harm data likely to be found in cohort studies.<sup>5</sup> Glazener and Evans simply described different outcomes that were reported in a few primary studies, such as alarm malfunction, false alarms, fright, failure to awaken the child, and awakening others.<sup>4</sup> Costs were not reported. This review did not incorporate an economic evaluation, and such information would contribute to decision making.

#### What are my patient's values and preferences for both the outcome we are trying to prevent and the side effects that may arise?

Each family will have to decide if the potential benefits of dry nights (eg, child's positive feelings about himself and reduced bed changes and laundry) are worth the cost of the alarm and the potential effects on loss of sleep of the child and the rest of the family. The clinician can recommend the use of alarms as a potential benefit, but informal cost-benefit considerations will have to be done by the family to make a decision.

#### Resolution of the scenario

Before answering the question in the scenario, we must now ask ourselves, "Is this a good review?" We can answer "yes" based on the criteria summarised in the box. The Glazener and Evans review addressed clear, focused questions; did a fairly extensive search; included randomised trials; applied predefined inclusion, exclusion, and validity criteria; and did a meta-analysis with RRs and odds ratios for most of the main outcomes.<sup>4</sup> It is a high quality review that found that alarms reduce nocturnal enuresis in children. As a clinician, you could feel quite confident in letting families know that alarms have the potential to help reduce enuresis, while informing them of the potential negative outcomes (costs, failed alarms, sleep interruptions) with less certainty. This guide applied the criteria of Sackett *et al*<sup>6</sup> to a specific example of an existing review about a real clinical problem. Application of the criteria can provide a useful analysis to decide if the results of a systematic review can be confidently used in practice.

#### Books

McKibbin A, Eady A, Marks S. *PDQ evidence-based principles and practice*. Hamilton: BC Decker, 1999.

Guyatt G, Rennie D, editors. *Users' guides to the medical literature. A manual for evidence-based clinical practice*. Chicago: AMA Press, 2001.

#### Articles

Greenhalgh T. Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ* 1997;**315**:672-5.

Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI How to use an overview. *JAMA* 1994;**272**:1367-71.

#### Online resources

Users' guides: on [www.cche.net/principles/content\\_all.asp](http://www.cche.net/principles/content_all.asp)  
BMJ series: [www.bmj.com/cgi/content/full/315/7109/672](http://www.bmj.com/cgi/content/full/315/7109/672)

*Useful resources for learning about critical appraisal of systematic reviews*

DONNA CILISKA, RN, PhD  
School of Nursing  
Faculty of Health Sciences  
McMaster University  
Hamilton, Ontario, Canada

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

SUSAN MARKS, BA, BEd  
Health Information Research Unit  
McMaster University,  
Hamilton, Ontario, Canada

- 1 Cullum N. Evaluation of studies of treatment or prevention interventions [editorial]. *Evidence-Based Nursing* 2000;**3**:100–2.
- 2 Cullum N. Evaluation of studies of treatment or prevention interventions. Part 2: applying the results of studies to your patients [editorial]. *Evidence-Based Nursing* 2001;**4**:7–8.
- 3 Sackett DL, Strauss SE, Richardson WS, et al. *Evidence-based medicine. How to practice and teach EBM*. Second edition. London: Churchill Livingstone, 2000.
- 4 Glazener CMA, Evans JHC. Alarm interventions for nocturnal enuresis in children. *Cochrane Database Syst Rev* 2001;(1):CD002911.
- 5 Roberts J, DiCenso A. Identifying the best research design to fit the question. Part I: quantitative designs [editorial]. *Evidence-Based Nursing* 1999;**2**:4–6.
- 6 Egger M, Smith G. Bias in location and selection of studies. *BMJ* 1998;**316**:61–6.
- 7 Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;**317**:1185–90.
- 8 Sheldon T. Statistics for evidence-based nursing [editorial]. *Evidence-Based Nursing* 2000;**3**:4–6.
- 9 Sheldon T. Estimating treatment effects: real or the result of chance? [editorial]. *Evidence-Based Nursing* 2000;**3**:36–9.
- 10 DiCenso A. Clinically useful measures of the effects of treatment [editorial]. *Evidence-Based Nursing* 2001;**4**:36–9.
- 11 Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses—sometimes informative, usually misleading. *BMJ* 1999;**318**:1548–51.

## Other articles noted

A list of the journals that are reviewed and the criteria for selecting articles from these journals for inclusion in *Evidence-Based Nursing* are included in each issue. In the past, the October issue of the journal has included a list of articles that passed all criteria but were not abstracted because, in the judgment of the editors, their findings were less widely applicable to nursing practice. As of the October 2001 issue of *Evidence-Based Nursing*, this list will be published on an ongoing basis on the EBN website only ([www.evidencebasednursing.com](http://www.evidencebasednursing.com)).

## Acknowledgements

As our fourth year of publication draws to a close, we would like to thank the many people who have contributed their time, expertise, and enthusiasm to *Evidence-Based Nursing*.

To our panel of advisers on qualitative research, who helped to select qualitative research papers, edited abstracts and commentaries for selected files, and who continue to help us to develop our approach to abstracting qualitative research:

Caroline Carlisle, Liverpool, UK  
 Marlene Z Cohen, Houston, Texas, USA  
 Jenny Ploeg, Hamilton, Ontario, Canada

Rita Schreiber, Victoria, British Columbia, Canada  
 Carl Thompson, York, UK  
 Sally Thorne, British Columbia, Canada

To the individuals who helped us in the selection of quantitative studies that were clinically important and relevant to practising nurses and who edited abstracts and commentaries for selected files\*:

\*Heather Arthur, Hamilton, Ontario, Canada  
 \*Peter Griffiths, London, UK  
 \*Sarah Hayward, Edmonton, Alberta, Canada  
 \*Andrew Jull, Auckland, New Zealand  
 \*Jeannette LeGris, Hamilton, Ontario, Canada

Brenda Morgan, London, Ontario, Canada  
 \*Janet Pinelli, Hamilton, Ontario, Canada  
 \*Jenny Ploeg, Hamilton, Ontario, Canada  
 Pauline Raynor, York, UK  
 Jennifer Skelly, Hamilton, Ontario, Canada

To Brian Hutchison and Trevor Sheldon, who provided thoughtful methodological advice about selected studies.

And finally, to the people who served as peer reviewers for the EBN notebook and implementation forum, for their thoughtful and timely feedback:

Kate Flemming, York, UK  
 Dorothy McCaughan, York, UK  
 Andrea Nelson, York, UK

Joyce Prince, Banbury, UK  
 Anne-Marie Rafferty, London, UK  
 Carl Thompson, York, UK