METHODOLOGY

Systematic Reviews: The Good, the Bad, and the Ugly

Yuhong Yuan, MD, PhD[1] and Richard H. Hunt, MB, FRCP, FRCPC, FACG, AGAF[1]

Systematic reviews systematically evaluate and summarize current knowledge and have many advantages over narrative reviews. Meta-analyses provide a more reliable and enhanced precision of effect estimate than do individual studies. Systematic reviews are invaluable for defining the methods used in subsequent studies, but, as retrospective research projects, they are subject to bias. Rigorous research methods are essential, and the quality depends on the extent to which scientific review methods are used. Systematic reviews can be misleading, unhelpful, or even harmful when data are inappropriately handled; meta-analyses can be misused when the difference between a patient seen in the clinic and those included in the meta-analysis is not considered. Furthermore, systematic reviews cannot answer all clinically relevant questions, and their conclusions may be difficult to incorporate into practice. They should be reviewed on an ongoing basis. As clinicians, we need proper methodological training to perform good systematic reviews and must ask the appropriate questions before we can properly interpret such a review and apply its conclusions to our patients. This paper aims to assist in the reading of a systematic review.

Introduction

Systematic review is crucial to the practice of evidence-based medicine. Systematic reviews can be qualitative, when the results of primary studies are summarized but not statistically combined, or quantitative, also called meta-analyses, when the results of primary studies are aggregated and statistical methods are used.

Rigorous research methods must be used to perform a systematic review, and strict rules apply to each step for generating the validated, necessary evidence: clinically relevant questions should be formulated, the systematic review should be carefully planned as for any other research project with a detailed protocol, eligibility criteria should be defined *a priori*, and search procedures must be comprehensive to identify all relevant studies (**Table 1**) (1,2). When meta-analysis is possible, appropriate methods should be used for data extraction, data combination, and analysis. Assessment for heterogeneity between studies is an important step, and subgroup analyses and sensitivity analyses should be used when necessary to assess the robustness of combined estimates (2,3). With proper reporting, others who read the review should reach the same summary of findings that the authors did.

Systematic reviews have strength of authority compared with narrative reviews. A well-conducted systematic review helps to answer specific, often narrow, clinical questions in depth, with comprehensive sources of information, criteria-based selection of patients and/or trials, and critical appraisal and synthesis of all relevant studies, and is thus usually evidence-based. In contrast, a narrative review usually addresses a broad range of issues related to a topic without specific literature sources, cites the literature selectively, mixes evidence with opinion, and often provides a qualitative summary and therefore has more potential for bias and is less likely to be evidence-based (**Table 2**) (4). "Expert opinions" reflected in narrative reviews may conclude with recommendations that are inconsistent with those of other experts or with the literature.

The good

Systematic reviews systematically evaluate and summarize current knowledge and help us to keep up to date when overwhelmed by the volume of medical literature. Many clinicians are interested in a meta-analysis of randomized controlled trials (RCTs) that investigate the effectiveness of a single intervention, and selection criteria are used to include RCTs that address the same question. Meta-analysis applies strict methods to combine data to provide a more reliable and enhanced precision of effect estimate than those achieved in individual studies. Meta-analysis can also answer some uncertainties, especially when individual studies have too small a sample size to prove a significant treatment effect or differing studies give controversial results. Thus, meta-analysis can reduce the probability of false-negative or false-positive results and potentially lead to a more timely introduction of effective treatments (5).

[1]Division of Gastroenterology, Department of Medicine, McMaster University Health Science Centre, Hamilton, Ontario, Canada. **Correspondence**: Richard H. Hunt, MB, FRCP, Division of Gastroenterology, Room 4W8A, Department of Medicine, McMaster University Health Science Centre, 1200 Main Street West, Hamilton, Ontario L8N 3Z5, Canada. E-mail: huntr@mcmaster.ca

Systematic reviews are not restricted to RCTs. Various designs of high-quality trials can be used to support limited information used in meta-analysis without recourse to low-quality trials that might be subject to bias (6). Observational case–control or cohort studies are commonly explored for adverse effects and for any association between risk factors and diseases or when questions could not be answered by RCTs; case reports and basic-science studies may also provide invaluable evidence. Systematic reviews provide an important evaluation of differing sources of evidence or aspects of a disease; these can include, for example, prevalence, etiology, screening and diagnostic testing, treatment efficacy and adverse effects, cost-effectiveness, prognosis, and prevention. Sometimes, several diseases are included in the same systematic review (e.g., a review of the adverse effects of a proton pump inhibitor (PPI) requires a review of all eligible studies, including those pertaining to treatment of peptic ulcer and gastroesophageal reflux); in other cases both RCTs and observational studies are required in the same systematic review (e.g., assessing the overall risk of myocardial infarction with nonsteroidal anti-inflammatory drugs (NSAIDs)) (7).

Systematic reviews are invaluable when applying for research funding because they help to establish what we know and don't know. This helps us to define the methods that should be used in future studies, such as justification of the sample size (8), and to establish proper eligibility of study participants who will benefit most from the study intervention. The summarized data provided by a meta-analysis can be further used, for example, in cost-effectiveness analysis. Thus, systematic reviews and meta-analyses provide the highest levels of evidence to guide clinical decision making and support practice guidelines.

There are good examples of systematic reviews that have contributed to the practice of gastroenterology. The effect of *Helicobacter pylori* eradication was unclear in early small trials until meta-analyses combined results with enough power to confirm the best pooled eradication rates (9). Subsequent cost-effectiveness meta-analyses provided economic evidence and further supported *H. pylori* eradication as first-line therapy for patients with duodenal ulcer (10). A cost-effectiveness analysis using data from a meta-analysis concluded in the early 2000s that 7-day PPI triple therapy is the most cost-effective strategy (11). Because of increasing eradication failure seen with triple therapy, a sequential therapy was recently proposed, and a meta-analysis has suggested that this approach is superior to standard triple therapy for *H. pylori* eradication in patients naive to treatment (12). The evolution of *H. pylori* eradication therapy has recently been reviewed by Gisbert *et al.*, who undertook a formal review of meta-analyses of eradication regimens. They concluded that the ideal future therapy remains to be designed (13), thus indicating the importance of an ongoing review of evidence. The synergy between *H. pylori* infection and NSAIDs was quantified by the publication of a meta-analysis that indicated that the risk of ulcer in *H. pylori*–infected patients taking NSAIDs is some 61-fold higher than that in *H. pylori*–negative patients not taking NSAIDs (14).

Table 1. Basic principles of conducting a systematic review

Formulate the research question(s)
Conduct the literature search
Specify all selection and assessment methods
Detail the procedure for data extraction
Provide the approach to analysis
Adapted from ref. 2.

Table 2. Differences between systematic reviews and narrative reviews

Feature	Systematic review	Narrative review
Question	Often a focused clinical question	Often broad in scope
Sources and search	Comprehensive sources and explicit search strategy	Not usually specified, potentially biased
Selection	Criterion-based selection, uniformly applied	Not usually specified, potentially biased
Appraisal	Rigorous critical appraisal	Variable
Synthesis	Qualitative summary that includes statistical synthesis (meta-analysis)	Often a qualitative summary
Inferences	Usually evidence-based	Sometimes evidence-based
Adapted from ref. 4.		

Applying a different approach of two separate meta-analyses of both pharmacodynamic pH data and clinical trial data, we were able to define the three primary determinants for the healing of peptic ulcer and gastroesophageal reflux disease with acid-suppression treatments as the degree and duration of acid suppression over 24 hours and the duration of treatment (15–18). This evidence supported the rationale of developing more effective acid-suppressing drugs for acid-related disorders. Good systematic reviews can find out more about the disease area by exploring reasons for heterogeneity. For example, the Cochrane PPI and non-ulcer dyspepsia (NUD) systematic review showed that PPIs have different efficacy in different dyspepsia subgroups; PPI treatment was more effective in patients with reflux symptoms than in those with epigastric pain or dysmotility (19).

The bad

Systematic reviews can be subject to biases and can be flawed. Systematic reviews and meta-analyses are research projects by nature but retrospective in study design, and their quality is therefore directly related to that of the original studies. Bias can be induced in any step of meta-analysis, leading investigators to the wrong conclusions (20). Any problems that threaten the validity of the individual studies, especially bias and confound-

METHODOLOGY

ing, affect meta-analysis of such studies (e.g., observational studies) (5). Scientific principles and rigorous process are required to minimize any bias, and the quality of a systematic review depends on the extent to which scientific review methods are used to ensure exclusion or at least minimize bias.

Meta-analyses may have obvious methodological flaws such as the lack of a protocol or of adherence to it, or inappropriate meta-analysis techniques (21). A clinically relevant research question should be posed. Rigorous methodology can identify all relevant studies that address the research question, but if after the meta-analysis is started the question is changed or the inclusion criteria are expanded for whatever reason (e.g., the retrieved studies may not be enough to answer the question), the original search strategies may no longer be appropriate. When a systematic review is performed, the search must be wide enough to ensure retrieval of the most relevant information, but it is also possible to cast the net too wide, resulting in misclassification (22). It is common to see critical letters following publication of meta-analyses, when serious errors have been found that impair both internal and external validity (23).

A systematic review can suffer from the "garbage in, garbage out" phenomenon. The quality of trials is of crucial importance for the estimation of treatment effect. For instance, the magnitudes of benefit of a prokinetic and H_2RAs for NUD were higher in low-quality trials than in trials of high quality in previous meta-analyses, and so their treatment effect was overestimated (24). Many original studies underreported their results and the hidden information affected study eligibility. In others, the study results were negative, or the quality of the study may have been poor. In contrast, an original study may be misleading, unhelpful, or even harmful, even when it seems to be perfectly well designed, conducted, analyzed, and reported in a good journal (25). A series of biases may threaten the validity of clinical trials, including selection bias, performance bias, detection bias, and attrition bias (5). If a meta-analysis has included studies of low quality, then the conclusion of the meta-analysis will be open to question, meeting the criticism of "garbage in, garbage out." For example, an early meta-analysis critically appraised outcome measures in treatment trials of *H. pylori*–positive NUD patients and found significant methodological problems in all seven of the retrieved studies (26). These included variations in symptom measurement, assessment of dyspepsia severity, use of global assessments, and methods used to determine changes in symptom severity. Most trials on NUD and *H. pylori* infection have not used validated symptom measurement, and the authors suggested that consensus should be reached by investigators in this area (26). However, when we looked into this question again some 14 years later, there was still considerable variation in defining symptom improvement and in observation time points in the 39 included NUD RCTs. We reaffirmed the concerns that standardized objective outcome measures are still needed for NUD trials in the future (27). Simply combining trials without carefully looking into the variation in outcome measures will lead to wrong treatment-effect estimates and consequently to misleading conclusions.

Quality assessment has an impact on the effect estimate. Quality assessment is important for valid and clinically relevant effect estimates in a systematic review (28). Some meta-analyses draw conclusions based only on "high-quality" studies. However, given that more than 100 quality scales or modified scales have been used in different meta-analyses for assessing quality of included RCTs, there will be problems if we do not consider this issue carefully (29). Individual quality measures are not reliably associated with the strength of treatment effect, and different summary quality scales can yield different findings for the same research information (30,31). In some cases, the quality assessment should be tailored to include key methodological criteria that are important to the validity and interpretation according to the type of study (2). Several quality components for the assessment and control of bias have been suggested, but their effect on the extent and direction of bias requires evaluation. A number of studies suggest that lack of adequate randomization, unconcealed allocation, and lack of double-blinding are associated with an overestimate of treatment effect. A study in which workers from the Cochrane Collaboration replicated each of 70 meta-analyses found that two-thirds of conclusions that favored one of the interventions were no longer supported if only trials with adequate concealment of allocation were included (32). Double-blinding may not be applicable for certain trials, when one or more parties are difficult to blind, or where blinding fails because the active treatment and placebo were not sufficiently similar, or when some RCTs do not provide the necessary details regarding the blinding methods used. The influence of placebo might depend on the characteristics of the outcome measure; the type of outcome may be equally important, and measurable clinical outcomes such as mortality may be less prone to assessment bias than a subjective outcome such as pain (20). Finally, some biases are difficult to detect and appraise by meta-analysis, and the extent of bias in individual trials is always going to be somewhat unpredictable.

"Mixing apples and oranges" and heterogeneity detection. In systematic reviews, the results from different studies are always combined, and so heterogeneity raises an important concern that a meta-analysis has "mixed apples and oranges." Patient characteristics, disease condition, enrollment criteria, intervention, and outcome assessments are diverse across studies. When reading the conclusions from a systematic review, one should systematically evaluate the individual study methodology and patient characteristics and compare them with those we see in the clinic rather than interpret the overall pooled results. Obvious clinical heterogeneity should not be ignored, even if the statistical tests do not detect a significant heterogeneity between the studies. Clinical common sense and careful appraisal of the raw data help to detect clinical heterogeneity. However, it is common for a physician searching for results of a meta-analysis to apply

the conclusions in the clinic without considering the possible differences between his or her patient and those enrolled in the studies (especially when the meta-analysis presents a convincing forest plot with a brief but strong conclusion!).

Although subgroup analyses are important for interpreting statistical heterogeneity, too many subgroup analyses can induce a significant effect of chance. Sometimes, we just don't have the results for the subset in which we are particularly interested (e.g., elderly patients), or the significant heterogeneity observed cannot be eliminated, even after several subgroup analyses are tried. In a meta-analysis of levofloxacin-based rescue regimens (LBRs) for *H. pylori* eradication, the pooled estimate changed from nonsignificant to significant in favor of LBRs when a single outlier study was excluded (the only study that significantly favored a quadruple regimen over LBRs), whereas the heterogeneity "markedly decreased" but was still significant (from $P < 0.0001$ to $P = 0.01$) (33). So, would this "significantly favored" conclusion apply to your patient? Considering the remaining significant heterogeneity, and, in particular, that a sensitivity analysis excluding extreme studies did not support the study conclusion, we should consider the effect of clinical and population differences before applying the conclusion to an individual patient. The generalizability of this meta-analysis is also limited given that 5 of 10 studies, including the 2 that most favored levofloxacin, came from the same investigators (34). More clinical trials from different centers and an updated meta-analysis should be performed to provide us with a robust result. Meta-regression is another useful method to explain heterogeneity at the study level, but this should also be interpreted with caution (35). For example, the relationship between the effect estimate and average patient characteristics across trials may not be the same as that relationship within trials (36). Subgroup analyses and meta-regression should be planned *a priori* to reduce analysis bias and allow an adequately powered analysis, but many meta-analyses have not done this. In fact, if the data handling/analysis methods are properly applied, we can still have a satisfactory mix of apples and oranges (35).

Publication bias has an impact on the validity of a systematic review. Studies without statistical significance (negative studies) or of small sample size are less likely to be published, and positive studies are more likely to result in multiple publications. Moreover, publication of studies is often selective on the basis of the direction and magnitude of their results. Therefore, systematic reviews cannot present the truth as completely as we believe or would wish. Pooling the results from published studies alone may lead to an overestimate of the effectiveness of an intervention. Publication bias can be introduced at many different steps. Through the application of proper techniques, publication bias and small-study effects can be detected (e.g., funnel plot) and minimized (e.g., inclusion of unpublished data) in meta-analyses, and they can best be prevented by prospective study registration with accessible results (37). However, it is impossible to identify all the "gray" literature. For instance, many systematic reviews include only English-language studies; as for systematic reviews that state that they "included all-language RCTs," the lack of a language restriction in a given database does not mean that studies in all languages are included (38). Other forms of reporting bias include time-lag bias, citation bias, and outcome-reporting bias (5). Clinicians using systematic reviews to guide their practice should remain aware of the dangers of these possible reporting biases.

Inappropriate handling of data can lead to wrong conclusions. First, missing data potentially introduce bias, especially when handled inappropriately. Assuming that dropouts were treatment failures or excluding dropouts from the analysis can lead to different effect-size estimates. Second, some systematic reviews call themselves a "pooled analysis" and add the numbers of events observed in a treatment group across trials and then divide the results by the total number of patients in the group without applying proper meta-analysis techniques. This ignores the randomization process in individual studies and can lead to wrong interpretations and conclusions (39). Third, inappropriate handling of data introduces further bias. For instance, a recent meta-analysis of the effect of hemoclipping for upper gastrointestinal bleeding included two studies that originally randomized patients into injection or mechanical-intervention groups (hemoclip or banding) (40). The choice of mechanical methods was arbitrary, and the patients were randomized neither to hemoclip vs. controls nor to hemoclip vs. banding. However, the authors took the data from the hemoclip subgroup, combined this with results from six other RCTs, and compared the total with the injection group; this has the potential to imbalance and thus bias against the benefits of randomization (41). The authors subsequently argued (42) that even if the "suggested" (proper) methods were used to handle the data, "the difference between the two groups remained unchanged" (although the magnitude of the difference did change), and that the difference in one subgroup became "marginally insignificant" (although it therefore no longer supported their original conclusion). We believe that inappropriate handling of data may be difficult to detect but will result in invalid results and lead to erroneous conclusions.

Sample-size consideration: the absence of evidence of a difference vs. evidence of the absence of a difference. The outcome of a meta-analysis is influenced by the inclusion or exclusion of certain trials and the degree of adherence to rigorous eligibility criteria. There are no rules regarding sample-size requirements for a meta-analysis. Most believe that, to warrant confidence that the conclusion is not a false negative due to an insufficient number of patients, the total number of patients included in a meta-analysis should be at least as large as that in a well-designed and optimally powered RCT that is able to detect a significant difference, and, considering heterogeneity, a larger total number of patients may be needed (21). On the other hand, some authors believe that thousands of patients are not needed to show a statistically significant difference, the clinical implication of which is unclear, whereas others argue that an under-

powered sample size will lead investigators to "mistake the absence of evidence of a difference for evidence of the absence of a difference." For example, two meta-analyses showed that endoscopic dual therapy is more efficacious than injection alone for patients with high-risk peptic ulcer bleeding (43,44), but the second meta-analysis suggested that dual therapy had no advantage over thermal or mechanical monotherapy, on the basis of two subgroup analyses that included only three studies each (n = ~70–140 per arm) (44). This sparked two arguments. One was that "the power is too low to draw that conclusion (0.4 and 0.2 respectively)" (45). The other, whose authors asked, "Do we need a sample size of over 6,000 patients to show a significant difference between dual therapy and mechanical monotherapy?," noted that "the evidence of the absence of a difference is the best evidence to date" (46). As of now, there is no correct answer, and sometimes we should judge the clinical significance of the difference rather than the statistical significance.

The ugly

Meta-analysis can be misleading. Systematic reviews of questionable validity can be misleading and have the potential for an adverse impact on clinical practice. Recently, a group of experts who identified and replicated eight meta-analyses that addressed drug therapy of irritable bowel syndrome found many errors in both the application of eligibility criteria and dichotomous data extraction in all eight meta-analyses. There were errors in 15 (94%) of 16 reported pooled treatment effects; these were overestimated in 10, underestimated in 5 and remained the same in only 1 analysis (A. Ford, G.H. Guyatt, N.J. Talley *et al.*, personal communication). The impact of errors was substantial, with a ≥10% change in the relative pooled treatment effect in five meta-analyses when it was recalculated; in another four meta-analyses, the statistical significance of the pooled treatment effect also changed (A. Ford, G.H. Guyatt, N.J. Talley *et al.*, personal communication). Furthermore, simply combining different diagnoses of a disease but applying a general conclusion to all is misleading. In the clipping meta-analysis discussed earlier (40), the authors chose to combine Dieulafoy lesion studies and peptic ulcer bleeding studies despite the unique endoscopic and histopathological characteristics of Dieulafoy lesions, which are more likely to respond to mechanical endoscopic hemostasis than is peptic ulcer disease. Nevertheless, the authors drew the same conclusion as for nonvariceal upper gastrointestinal bleeding as a whole. We disagree, and believe that the conclusions of a meta-analysis conducted without subgroup analysis should not be applied to all (47).

Meta-analyses can be misused. Even when a meta-analysis is performed correctly, it can be misinterpreted and have an adverse impact on clinical practice. A research question about an intervention has four key components: population, condition, intervention, and outcome. Applying the conclusions of a meta-analysis to aid clinical decision making requires the clinician to compare an individual patient with patients in the meta-analysis for the four components. Obviously, conclusions from a meta-analysis of studies in the elderly may not apply to a young patient; the superiority of a new drug drawn from placebo-controlled trials does not mean that the intervention is superior to standard treatment; a better efficacy in one outcome does not mean it will be equivalent for all outcomes. For example, a meta-analysis for peptic ulcer bleeding suggests that oral or intravenous PPIs reduce rebleeding and the need for surgery and repeat endoscopic treatment as compared with placebo or H_2RA, but there is no overall reduction in all-cause mortality, although PPIs improve mortality among Asian patients and those at highest risk (48). Therefore, we have no evidence to support the efficacy of PPIs with respect to mortality in non-Asian or low-risk patients. Similarly, a current meta-analysis of first-line, PPI-based triple therapy for *H. pylori* eradication concluded that "available data suggest that extending triple therapy beyond 7 days is unlikely to be a clinically useful strategy" (49), which may give the impression that 7 days of treatment is long enough. However, only 4 of the 21 included studies were of high quality, only 3 clearly described concealed allocation, and only 4 were double-blind. Subgroup analyses of 7-day vs. 14-day regimens showed a significant difference for amoxicillin-containing therapy but no significant difference in the subgroup of metronidazole-containing therapies. Considering the limited data from this meta-analysis, and the facts that many studies contain old data and that there have been increasing clarithromycin resistance and increasing reports of treatment failure in recent years, simply applying the conclusion of this meta-analysis and prescribing only 7 days of initial triple therapy to each patient may not be enough, especially in regions with high resistance rates.

Systematic reviews usually focus on one side of the evidence and cannot answer all clinically relevant questions. Some clinicians prefer narrative reviews to systematic reviews because the latter ask a question that is too "narrow" for the clinical scenario. Systematic reviews that explore the evidence of harm, as well as benefit, are essential for decision making, but sometimes harm is not considered at all in systematic reviews. For example, when the GRADE Working Group developed the new system to grade the quality of evidence and the strength of recommendations, much of the information they found to be lacking was missing in the original systematic reviews, particularly information about harms and side effects (50). Applying evidence for benefit without considering the possible harms of an intervention may result in more harm than good. It is also common that the exact conditions and clinical indications are not clearly described in a systematic review, so that the reader cannot be sure whether the conclusions apply to an individual patient. Sometimes, the question we ask is not answered by published meta-analyses, and then we may misuse a conclusion from a meta-analysis that asked a somewhat similar question, without considering to what extent it is appropriate for our patient.

Systematic reviews may be difficult to incorporate into practice. It is very common for global or regional guidelines to refer to systematic reviews, especially meta-analyses, to support the

recommendations. Guidelines have potential benefits but may also have limitations that could even lead to harm (51–53). It is agreed that high-quality evidence need not result in a strong recommendation. It has been argued that systematic errors in meta-analysis could influence the final recommendations of guidelines (23), and guidelines based on level 1 evidence, which ignores non-RCT research, can also result in unnecessary death (54). Studies included in a meta-analysis might have been performed in a limited number of patients or a particular geographic region, and so the conclusions may not be applicable to patients globally. For example, the conclusion of a meta-analysis about upper gastrointestinal bleeding that included mostly Asian trials may not apply to a Western country. Even if the recommended intervention is "ideal," it may be difficult or impossible to implement in some clinics, and some medications are too expensive or are unavailable in developing countries. If the efficacy of a new treatment is only marginally better than that of an existing treatment by meta-analysis, it may not be worth changing current practice in some geographic regions. It might be better to maintain an available, inexpensive intervention that works locally, rather than adopt a new and costly intervention that is difficult to implement, needs time and effort to train clinicians, and—where incremental benefit is marginal, or even in local hands—not as good as the original treatment. Furthermore, when "no significant difference" is reported between two interventions in a systematic review, we should always ask "the difference between what and what?" and consider what kind of difference interests us in clinical practice.

Commentary

Dramatic advances in both basic and applied research have increasingly influenced the practice of gastroenterology. Compared with other specialists, gastroenterologists use a wider range of diagnostic tests, work with colleagues from a broader range of disciplines, rely on many small trials of therapy, are faced with more chronic relapsing conditions, and use techniques that are more operator-dependent. It is therefore a challenge to practice evidence-based medicine in gastroenterology (55).

A systematic review is a scientific research tool that is valuable in clinical practice and has many benefits compared with a narrative review. However, systematic reviews are also subject to various types of bias and can be flawed. To perform a proper systematic review and meta-analysis, we should ask the right and relevant question, have a proper protocol that is adhered to, use rigorous methodology to perform the literature search and data analysis, investigate the statistical as well as the clinical heterogeneity, and use biological and clinical sense to interpret the conclusions. Physicians performing systematic reviews should have proper methodological training to minimize bias and avoid flaws.

Instead of performing systematic reviews on their own, many physicians look for published systematic reviews, especially meta-analyses, to answer the clinical questions encountered in their practice. When a practicing clinician interprets a systematic review, the strength of evidence, the methodology, and

Table 3. Some questions to ask when interpreting the results of a systematic review
1. Do the objective and the inclusion criteria for this review match the question to which I am seeking an answer?
2. Does the review state the search methods that were used to find studies relevant to the main questions of the review?
3. Was the search for studies reasonably comprehensive?
4. How likely is it that additional relevant research has been started, completed, or published since this review was done?
5. Are the criteria used for deciding which studies to include in the review reported?
6. Did the reviewers avoid bias when deciding whether studies were eligible?
7. Did the selection of studies (in particular, the choice of eligible study designs) minimize the possibility of including studies with a high propensity for bias?
8. Were appropriate criteria used to assess the validity of the included studies?
9. Was the validity of all included studies assessed using these criteria, and were the findings reproducible?
10. Were appropriate methods used to combine the findings of the relevant studies?
11. Are the conclusions made by the authors supported by the results of the systematic review?
12. How likely is it that any additional relevant research that has been started, completed, or published since this review was done would change its conclusions?
13. Is there any reason to believe that the findings of this review are not relevant to the type of patient, intervention, or setting I am interested in?
Adapted from ref. 56.

the validity should be carefully considered. Recently, Michael Clarke of the Cochrane Collaboration provided a very useful table listing the questions we should ask when interpreting the results of such a review (Table 3). Accurate interpretation will help a clinician decide the relevance of the review and what it means in terms of making informed clinical decisions in practice (56). Furthermore, no evidence should be considered final, and all systematic reviews should be subject to ongoing review.

Finally, available systematic reviews do not always fit the clinical question being asked; the answers most sought at the bedside are not likely to emerge from meta-analyses, nor from individual clinical trials. When using the best products of evidence-based medicine—meta-analysis or guidelines—we should always remember that evidence alone is never sufficient to make our clinical decisions; rather, the best evidence, experience, and patients' values are three core aspects according to which management decisions should be individualized.

CONFLICT OF INTEREST

Guarantor of the article: Richard H. Hunt.
Specific author contributions: The authors contributed equally.

METHODOLOGY

REFERENCES

1. Sackett DL, Rosenberg WM, Gray JA *et al.* Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71–2.
2. Haynes B. Conducting systematic reviews. In: Haynes RB, Sackett DL, Guyatt GH *et al.* (eds). Clinical Epidemiology: How to Do Clinical Practice Research, 3rd edn. Lippincott Williams & Wilkins: Philadelphia, 2006.
3. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. BMJ 1997;315:1533–7.
4. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997;126:376–80.
5. Zwahlen M, Renehan A, Egger M. Meta-analysis in medical research: potentials and limitations. Urol Oncol 2008;26:320–9.
6. Smith LA, Moore RA, McQuay HJ *et al.* Using evidence from different sources: an example using paracetamol 1000 mg plus codeine 60 mg. BMC Med Res Methodol 2001;1:1.
7. Scott PA, Kingsley GH, Smith CM *et al.* Non-steroidal anti-inflammatory drugs and myocardial infarctions: comparative systematic review of evidence from observational studies and randomised controlled trials. Ann Rheum Dis 2007;66:1296–304.
8. Sutton AJ, Cooper NJ, Jones DR *et al.* Evidence-based sample size calculations based upon updated meta-analysis. Stat Med 2007;26:2479–500.
9. Chiba N, Rao BV, Rademaker JW *et al.* Meta-analysis of the efficacy of antibiotic therapy in eradicating *Helicobacter pylori.* Am J Gastroenterol 1992;87:1716–27.
10. O'Brien B, Goeree R, Mohamed AH *et al.* Cost-effectiveness of *Helicobacter pylori* eradication for the long-term management of duodenal ulcer in Canada. Arch Intern Med 1995;155:1958–64.
11. Calvet X, Gené E, López T *et al.* What is the optimal length of proton pump inhibitor-based triple therapies for H. pylori? A cost-effectiveness analysis. Aliment Pharmacol Ther 2001;15:1067–76.
12. Jafri NS, Hornung CA, Howden CW. Meta-analysis: sequential therapy appears superior to standard therapy for *Helicobacter pylori* infection in patients naive to treatment. Ann Intern Med 2008;148:923–31.
13. Gisbert JP, Pajares R, Pajares JM. Evolution of *Helicobacter pylori* therapy from a meta-analytical perspective. Helicobacter 2007;12(Suppl 2):50–8.
14. Huang JQ, Sridhar S, Hunt RH. Role of *Helicobacter pylori* infection and non-steroidal anti-inflammatory drugs in peptic-ulcer disease: a meta–analysis. Lancet 2002;359:14–22.
15. Jones DB, Howden CW, Burget DW *et al.* Acid suppression in duodenal ulcer: a meta-analysis to define optimal dosing with antisecretory drugs. Gut 1987;28:1120–7.
16. Burget DW, Chiverton SG, Hunt RH. Is there an optimal degree of acid suppression for healing of duodenal ulcers? A model of the relationship between ulcer healing and acid suppression. Gastroenterology 1990;99:345–51.
17. Bell NJ, Burget D, Howden CW *et al.* Appropriate acid suppression for the management of gastro-oesophageal reflux disease. Digestion 1992;51 (Suppl 1):59–67.
18. Howden CW, Burget DW, Hunt RH. Appropriate acid suppression for optimal healing of duodenal ulcer and gastro-oesophageal reflux disease. Scand J Gastroenterol Suppl 1994;201:79–82.
19. Moayyedi P, Delaney BC, Vakil N *et al.* The efficacy of proton pump inhibitors in nonulcer dyspepsia: a systematic review and economic analysis. Gastroenterology 2004;127:1329–37.
20. Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006;163:493–501.
21. Verstraete M. Value and limitation of meta-analysis. Pathophysiol Haemost Thromb 2002;32:278–81.
22. Holmdahl L. A misleading meta-analysis of Seprafilm. World J Surg 2008; 32:1888–9.
23. Tobin MJ, Jubran A. Meta-analysis under the spotlight: focused on a meta-analysis of ventilator weaning. Crit Care Med 2008;36:1–7.
24. Abraham NS, Moayyedi P, Daniels B *et al.* Systematic review: the methodological quality of trials affects estimates of treatment efficacy in functional (non-ulcer) dyspepsia. Aliment Pharmacol Ther 2004;19:631–41.
25. Ioannidis JP. Perfect study, poor evidence: interpretation of biases preceding study design. Semin Hematol 2008;45:160–6.
26. Veldhuyzen van Zanten SJ. A systematic overview (meta-analysis) of outcome measures in *Helicobacter pylori* gastritis trials and functional dyspepsia. Scand J Gastroenterol Suppl 1993;199:40–3.
27. Wang CC, Yuan Y, Hunt RH. Effect of placebo in controlled trials of patients with functional dyspepsia: a meta-analysis. Gastroenterology 2007;132(Suppl 2):A-94. Abstract 645.
28. Verhagen AP, de Vet HC, de Bie RA *et al.* The art of quality assessment of RCTs included in systematic reviews. J Clin Epidemiol 2001;54:651–4.
29. Juni P, Witschi A, Bloch R *et al.* The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282:1054–60.
30. Balk EM, Bonis PA, Moskowitz H *et al.* Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002;287:2973–82.
31. Moyer A, Finney JW. Rating methodological quality: toward improved assessment and investigation. Account Res 2005;12:299–313.
32. Pildal J, Hróbjartsson A, Jørgensen KJ *et al.* Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol 2007;36:847–57.
33. Gisbert JP, Morena F. Systematic review and meta-analysis: levofloxacin-based rescue regimens after *Helicobacter pylori* treatment failure. Aliment Pharmacol Ther 2006;23:35–44.
34. Yuan Y, Thabane L, Hunt RH. Levofloxacin-based rescue regimens after *Helicobacter pylori* treatment failure: how strong is the evidence? Aliment Pharmacol Ther 2006;23:1283–5.
35. Moayyedi P. Meta-analysis: can we mix apples and oranges? Am J Gastroenterol 2004;99:2297–301.
36. Reade MC, Delaney A, Bailey MJ *et al.* Bench-to-bedside review: avoiding pitfalls in critical care meta-analysis—funnel plots, risk estimates, types of heterogeneity, baseline risk and the ecologic fallacy. Crit Care 2008;12:220.
37. Montori VM, Smieja M, Guyatt GH. Publication bias: a brief review for clinicians. Mayo Clin Proc 2000;75:1284–8.
38. Pilkington K, Boshnakova A, Clarke M *et al.* "No language restrictions" in database searches: what does this really mean? J Altern Complement Med 2005;11:205–7.
39. Lièvre M, Cucherat M, Leizorovicz A. Pooling, meta-analysis, and the evaluation of drug safety. Curr Control Trials Cardiovasc Med 2002;3:6.
40. Sung JJ, Tsoi KK, Lai LH *et al.* Endoscopic clipping versus injection and thermo-coagulation in the treatment of non-variceal upper gastrointestinal bleeding: a meta-analysis. Gut 2007;56:1364–73.
41. Yuan Y, Wang C, Hunt RH. Re-analysis of application of haemoclip for non-variceal upper gastrointestinal bleeding data. Gut 2008;57:1023–4.
42. Tsoi KK, Sung JJ, Lau JY. Authors' reply. Gut 2008;57:1024.
43. Calvet X, Vergara M, Brullet E *et al.* Addition of a second endoscopic treatment following epinephrine injection improves outcome in high-risk bleeding ulcers. Gastroenterology 2004;126:441–50.
44. Marmo R, Rotondano G, Piscopo R *et al.* Dual therapy versus monotherapy in the endoscopic treatment of high-risk bleeding ulcers: a meta-analysis of controlled trials. Am J Gastroenterol 2007;102:279–89.
45. Calvet X, Vergara M, Gisbert JP *et al.* Dual versus endoscopic monotherapy in bleeding peptic ulcers. Am J Gastroenterol 2007;102:1826–7.
46. Marmo R, Rotondano G, Bianco MA *et al.* Can the evidence be linked to the alpha error? Am J Gastroenterol 2007;102:1827–8.
47. Yuan Y, Wang C, Hunt RH. Endoscopic clipping for acute nonvariceal upper-GI bleeding: a meta-analysis and critical appraisal of randomized controlled trials. Gastrointest Endosc 2008;68:339–51.
48. Leontiadis GI, Sharma VK, Howden CW. Proton pump inhibitor therapy for peptic ulcer bleeding: Cochrane Collaboration meta-analysis of randomized controlled trials. Mayo Clin Proc 2007;82:286–96.
49. Fuccio L, Minardi ME, Zagari RM *et al.* Meta-analysis: duration of first-line proton-pump inhibitor based triple therapy for *Helicobacter pylori* eradication. Ann Intern Med 2007;147:553–62.
50. Atkins D, Briss PA, Eccles M *et al.* Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. BMC Health Serv Res 2005;5:25.
51. Woolf SH, Grol R, Hutchinson A *et al.* Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ 1999;318:527–30.
52. Fried M, Quigley EM, Hunt RH *et al.* Are global guidelines desirable, feasible and necessary? Nat Clin Pract Gastroenterol Hepatol 2008;5:2–3.
53. Fried M, Quigley EM, Hunt RH et al. Is an evidence-based approach to creating guidelines always the right one? Nat Clin Pract Gastroenterol Hepatol 2008;5:60–1.
54. Tobin MJ. Counterpoint: evidence-based medicine lacks a sound scientific base. Chest 2008;133:1071–4.
55. Rosenberg W. Introduction. In: Irvine EJ, Hunt RH (eds). Evidence-Based Gastroenterology. BC Decker: Hamilton, Ontario, Canada, 2001, pp 1–7.
56. Clarke M. Interpreting the results of systematic reviews. Semin Hematol 2008;45:176–80.