WJ350303_258-265.qxp 4/22/08 2:28 AM Page 258

J Wound Ostomy Continence Nurs. 2008;35(3):258–265.
Published by Lippincott Williams & Wilkins

SPOTLIGHT ON RESEARCH

From the Center for Clinical Investigation

Systematic Reviews and Meta-Analysis

Studies of Studies

Sandra Engberg

Systematic reviews are designed to answer a focused clinical question. They employ a predetermined explicit methodology to comprehensively search for, select, appraise, and analyze studies. Meta-analysis is the statistical pooling of the results of studies that are part of a systematic review. Systematic reviews are research studies and, like other studies, they need to be based on a structured and valid methodology and take measures to minimize bias. High-quality systematic reviews can be powerful tools to support clinical decision-making, as well as summarize current knowledge in relation to an area of research interest. This article describes the methodology that should be used when doing a systematic review, presents guidelines for reporting the review, and provides a guideline for critically appraising published reviews.

When searching for the best evidence to support or change your clinical practice, to write a clinical paper or to prepare a grant proposal, it is increasingly likely that there will be systematic reviews or meta-analyses among the results of your search. By systematically identifying, appraising, synthesizing, and, if appropriate, statistically combining studies on a specific topic, systematic reviews have the potential to provide high-quality research evidence to guide clinical practice and support research proposals.[1] Meta-analysis is the statistical pooling of the results of studies that are part of a systematic review.[2]

The increase in the number of systematic reviews has paralleled the evidence-based practice movement. A search of OVID MEDLINE from 1992 to 1996 using "systematic review" in the title yielded 61 articles. Between 1997 and 2002, that number increased to 1,617, while 4,859 articles were retrieved between 2003 and 2007. Busy clinicians often lack both the time and skills needed to find, review, critically appraise and synthesize the large number of studies that may be available on a clinical topic. When available, good systematic reviews can be invaluable tools in helping both clinicians and researchers determine the state of the science on a particular clinical problem. In classifications of levels of evidence, systematic reviews are included in the highest level of evidence.[3–7]

Narrative Reviews of the Literature

A narrative review of the literature, sometimes called an integrative review, is generally completed by someone with expertise in the topic being presented. The author selects the articles to be included, and while such reviews can be useful, there is evidence that they are sometimes of poor quality.[8] The sources searched and strategies used to identify and select studies for inclusion are often not specified and the extent to which they are critically appraised is variable.[9] In narrative reviews, authors often use subjective methods to select and interpret studies and tend to include and report findings that consciously or unconsciously support their biases and interests in relation to the topic.[8,10–13] Consequently, while such reviews are useful when seeking a broad perspective on a topic, they are less often useful in providing unbiased quantitative answers to specific clinical questions.[9]

Systematic Reviews

In contrast to a narrative review, a systematic review is designed to answer a focused clinical question and employs a predetermined explicit methodology to comprehensively search for, select, appraise, and analyze studies.[9,14] The scientific rigor of this process decreases bias and is what makes a systematic review research and distinguishes it from a narrative review.[15] Over the past 3 decades, there has been a significant growth in the science of systematic reviews including statistical techniques, production centers, software, and quality standards.[16] Like other research studies, a protocol is developed before beginning the review. Table 1 summarizes the steps in a systematic review. The methods used in each step should be reported so that the process used is transparent and replicable.[16]

Sandra Engberg, PhD, RN, CRNP, University of Pittsburgh School of Nursing, Pittsburgh, Pennsylvania.
Corresponding author: Sandra Engberg, PhD, RN, CRNP, 440 Victoria Building, University of Pittsburgh, 3500 Victoria Street, Pittsburgh, PA 15261 (sje1@pitt.edu).

Copyright © 2008 by the Wound, Ostomy and Continence Nurses Society

WJ350303_258-265.qxp 4/22/08 2:28 AM Page 259

TABLE 1.

Steps in a Systematic Review

1. Formulate the question to be addressed
2. Identify the criteria that will be used to select studies
3. Plan, implement, and disclose the literature search strategy
4. Identify relevant studies
5. Extract data from the selected studies
6. Analyze the extracted data
7. Draw conclusions based on the data

The Question

A clearly formulated question is central to a high-quality systematic review. The question will define which studies should be included, guide the search strategy used to identify relevant studies, and determine what data need to be extracted from each study. Counsell[17] recommends including 4 elements in the clinical questions addressed by systematic reviews: (1) the patients of interest, (2) the exposure of interest, (3) the comparison exposure, and (4) the outcome(s) of interest. Depending on the type of review, the exposures (of interest and comparison) may include specific interventions, risk or prognostic factors or diagnostic studies. In addition to the primary question of the review, the reviewer may poise secondary questions. These questions should also be stated a priori and form the basis for secondary analyses such as examining the efficacy on the intervention of interest in subgroups.[14]

In a published systematic review, the question is often reworded as the aim or purpose of the review. In their recently published protocol for a planned systematic review, Garcia and colleagues[18] identified their aim as "to assess the efficacy of care in ostomized patients by nurses with a specific training in stomatherapy versus the care by staff nurses without a specific training in stomatherapy."(p2) This purpose contains the 4 elements that Counsell[17] and Meade and Richardson[19] identify as a critical starting point for a high-quality review: (1) *the patients of interest:* ostomized patients, (2) *the intervention exposure of interest*: care provided by nurses with specific training in stomatherapy, (3) *the comparison intervention exposure*: care provided by staff nurses without specific training in stomatherapy, and (4) *the outcome of interest*: efficacy of care. The authors operationally define the *efficacy of care* as early and late stoma complication and quality of life.

Identifying Criteria for Selecting Studies

The second step in conducting systematic reviews is identifying the criteria that will be used to select studies for inclusion in the review. The goal is to ensure that as many relevant studies as possible are identified. The selection criteria should match the clinical question. The reviewer should compile a set of explicit inclusion criteria based on the elements of the review question.[19] The intervention (exposure) of interest may include not only the type of treatment or exposure but also the intensity, frequency, and duration, depending on the question of interest. The comparison (control) intervention needs to be specified as well and, depending on the type of review, may include another treatment or placebo, absence of a risk or prognostic factor, or a gold standard diagnostic test.[17] The outcomes of interest should be defined and translated into explicit selection criteria. The patient population should be specified and may include a specific disease or health problem, clinical setting, and/or patient characteristics. Systematic reviews of treatments should include both beneficial and adverse outcomes.[17] Inclusion criteria often also include the timeframe for measuring the outcome(s).

Additional inclusion criteria should also be specified. One of these is the languages in the studies that will be included. Since studies are published in different languages, selecting only English language articles may bias the results of the review.[19] Financial and time limitations may, however, preclude the inclusion of all studies regardless of language since this will generally mean that translation is needed. The reviewer should also specify the study designs that will be included. The types of studies that should be included will vary with the clinical question that the review is addressing. When the question asks about the efficacy of an intervention, ideally the review includes only randomized clinical trials using an intention-to-treat approach when analyzing efficacy and objective, ideally blinded, outcome measures. For many nursing interventions, however, there may be few studies meeting these criteria, and reviewers need to consider other study designs that could answer their questions. The timeframe for the review should also be specified.

Another a priori decision that the reviewer needs to make is whether or not to include gray literature. Gray literature includes unpublished studies and studies published in sources other than widely available journals.[20] Examples include conference abstracts, research reports, book chapters, dissertations, unpublished data, policy documents, and personal correspondence. There are advantages and disadvantages of including this type of literature in a systematic review. The primary concerns about this literature are that, unlike research reports published in peer-reviewed journals, these studies may have been either rejected by a journal after being peer reviewed or never submitted and, therefore, not subjected to peer review. Both raise concern about the quality of the studies. All unpublished studies are not, however, of low quality, and not all published studies are of high quality.[20] In 2 reviews, studies with significant results were significantly more likely to be published than those without significant results.[21,22] In the Dickersin and Min[21] study, statistical significance was the only study characteristic significantly associated with publication.

WJ350303_258-265.qxp 4/22/08 2:28 AM Page 260

TABLE 2.

Electronic Databases

Database	Covers
CINAHL	Nursing and allied health literature. References, journal articles, books, dissertations, and patient-education materials. May be available through university or hospital library system
MEDLINE	All health and biomedical sciences, including medicine, nursing, dental medicine, public health, pharmacy, and allied health. Widely available through university and hospital libraries
PubMed	All health and biomedical sciences, including medicine, nursing, dental medicine, public health, pharmacy, and allied health. Internet access to abstracts and some full-text articles is free at http://www.ncbi.nlm.nih.gov/pubmed
EMBASE	Covers pharmacological and biomedical literature with comprehensive international coverage. May be available through university or hospital library systems
PsycINFO	All fields of psychology and the psychological aspects of related disciplines including medicine, psychiatry, nursing, sociology, education, pharmacology, physiology, linguistics, anthropology, business, and law. May be available through university or hospital library systems
AMED	Articles related to complementary and alternative medicine as well as allied health. May be available through university or hospital library systems
Cochrane Library	Numerous systematic reviews prepared by members of the Cochrane Collaboration; reviews are updated regularly. Free Internet access to abstracts (http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME?CRETRY=1&SRETRY=0); access to full-text reviews may be available through university or hospital library systems
Cochrane Central Register of Controlled Clinical Trials	Bibliographic references to controlled trials in health care together with references to clinical trials identified by contributors to the Cochrane Collaboration in MEDLINE and EMBASE. May be available through university or hospital library systems
PEDro	Physiotherapy Evidence Database. Abstracts of published clinical trials and systematic reviews on topics of interest to physiotherapists (some are also of potential interest to wound, ostomy, and continence nurses). Internet access to abstracts with links to full-text versions of some articles (http://www.pedro.fhs.usyd.edu.au/)

Excluding gray literature may introduce publication bias and threaten the validity of a review.[20] Gerber and colleagues[23] compared the proportion of meta-analyses published in 4 general medical (*Annals of Internal Medicine, British Medical Journal, Journal of American Medical Association, and Lancet*) and 4 specialty (*American Journal of Cardiology, Cancer, Circulation,* and *Obstetrics and Gynecology*) journals between 1932 and 2002 that reported including gray literature in their search strategies. In 2001/2002, 51% reported searching for gray literature. McAuley and coworkers[24] examined the impact of excluding gray literature on the estimates of intervention effectiveness in a random sample of 135 meta-analyses published between 1966 and 1995. Excluding gray literature from the meta-analyses resulted in significantly larger estimates of treatment effectiveness. Hopewell and colleagues[25] conducted a systematic review of studies investigating the impact of including gray literature in randomized trials of interventions ($n = 5$ studies). On average, published trials showed a 9% greater treatment effect than trials from the gray literature. The findings from these 2 reviews suggest that failure to include gray literature in systematic reviews is likely to introduce bias into a review and may overestimate treatment effectiveness.

After determining the inclusion criteria for selecting studies, it is generally helpful to develop a checklist of the selection criteria. Using a form during the literature, search will simplify the selection process, increase reliability, and provide a record of judgments made in relation to each identified study.[19] Like all data collection instruments, the form should be pilot tested for clarity, ease of use, and reliability, and revised, if necessary, prior to use in the review.

Identifying Relevant Studies

Once the inclusion criteria are identified, the next step is to conduct the literature search. To meet the goal of identifying all relevant studies, multiple search strategies need to be employed. A simple MEDLINE search has been shown to identify only about half of the evidence available from clinical trials.[26] Multiple databases should be searched with the specific databases included based on the topic of the review. Table 2 provides a list of electronic databases: what they cover and how they can be accessed. Prior to starting the search, search terms (keywords) that reflect the inclusion criteria need to be identified. Different search terms may need to be used when searching different databases. A reference librarian is an invaluable resource in constructing and implementing the search strategy for a systematic review.

Even when more than 1 database is searched, potentially relevant studies may not be identified.[17] Additional strategies should be employed including a search of the Cochrane Central Register of controlled trials (Table 2)

WJ350303_258-265.qxp 4/22/08 2:28 AM Page 261

and reviewing the reference lists of the relevant published studies that were found.[26] Additional studies may be identified by hand researching journals relevant to the review topic. This is, however, time consuming. While current evidence supports including gray literature in reviews, particularly when examining the efficacy of interventions, accessing this literature can be challenging. Search strategies that can be used include examination of presentation abstracts, Internet searches, and contacting experts and colleagues in the field.

After pilot testing the inclusion criteria on a sample of studies, experts recommend that 2 reviewers apply them independently to select potentially relevant studies from those retrieved using the search strategies.[19,26] This is recommended to reduce the likelihood of bias and errors during the selection process.[27] A consensus method should be used to resolve any disagreements about the inclusion of studies, involving a third reviewer as necessary.[26]

Extracting Data From Selected Studies

Once the studies that will be included in the review are identified, the next step is to review the selected studies using a standardized form that has been pilot tested with known studies. During this process, studies are critically appraised to assess their quality (validity). In addition, the relevant data that will be part of the review are extracted from each study. The data extracted during the review will vary depending on the purpose of the review (the question addressed) but may include information about the characteristics of the subjects in each study, the setting, interventions, outcome measures, and follow-up.[15] At least 2 reviewers should independently review each paper to extract the data and evaluate the study quality, and the same method used to achieve consensus about the selection of studies should also be used during this phase of the review.[26] Some experts recommend blinding reviewers to the author, journal, and funding source during the quality-review process.[15] Jadad and associates[28] examined the impact that blinding reviewers had on quality ratings. Blinded assessments produced significantly lower and more consistent quality scores than unblinded evaluations.

Study quality refers to the extent to which the design and conduct of the study (including analysis of the data collected) minimized the potential for bias.[19] The criteria used to evaluate quality will vary with the focus of the review.[19] Randomized controlled trials (RCT) should be evaluated for 4 types of bias: (1) selection bias, (2) performance bias, (3) attrition bias, and (4) detection bias.[15,29] Selection bias is related to the method used to randomize subjects. Bias is reduced by concealment—preventing investigators from knowing to which group subjects will be assigned prior to randomization. The goal is to prevent anyone from influencing the allocation decision (ensuring that it is truly random). Performance bias is systematic differences in the care subjects receive other than the intervention being studied. The most effective way to prevent performance bias is to blind both subjects and those delivering the care to subjects' group assignment. Attrition bias refers to systematic differences between the treatment and control (comparison) groups resulting from the loss of subjects from the study. Detection bias refers to systematic differences in the evaluation of outcomes in the treatment and control groups. The most effective method for protecting against detection bias is for the person(s) assessing the outcomes to be blinded to the group to which subjects were assigned. The quality evaluation of RCTs should examine the extent to which measures were taken to prevent each type of bias. As the risk of bias increases (ie, more criteria for reducing the 4 threats are not met), confidence in the results of the study decreases (the validity of the findings is called into question). If the information presented in the paper is unclear, the reviewer should try to contact the author to provide clarification.[29]

When quasi-experimental studies (intervention studies where subjects are not randomly assigned to groups) are included in the review, there will be differences in the treatment and control groups that could affect the outcomes making an intervention appear to be more or less effective than it actually is. Investigators may have taken steps such as matching treatment and control subjects on characteristics that could affect the outcome or controlling for those characteristics when examining the outcomes to try to limit bias. It is impossible, however, to control for all characteristics that could affect the outcome (confounders) since some factors are invariably unknown or not evaluated.[29] Despite the concerns about the quality of quasi-experimental studies, reviews may include them if there are few or no RCTs available.

The quality assessment can be used in varying ways in the review. Sometimes it is used to exclude studies; studies below a prespecified quality level are excluded. It can also be used to explain inconsistencies in the results of studies reviewed. Some reviewers will elect to keep all studies and do a sensitivity analysis, to determine if the overall results are the same when the lower-quality studies are included in the analysis as when they are excluded. Finally, some reviewers will weight studies according to their quality when combining the results of the studies during data analysis. In this way, higher-quality studies have more influence of the summary result.[29] There is no agreement on which approach should be used.

Analyzing the Extracted Data

During this phase of the review, the findings of the individual studies included are synthesized to answer the question(s) asked. The analysis should start with a qualitative synthesis of the extracted data. This synthesis should include characteristics of the individual studies including,

WJ350303_258-265.qxp 4/22/08 2:28 AM Page 262

for example, subject demographic characteristics, characteristics (eg, severity) of the health problem and the intervention being studied, and characteristics of the studies. This synthesis will allow the reviewer to see the overall characteristics of the studies included. It can also identify substantial differences between the studies (clinical heterogeneity) that may make it inappropriate to synthesize the data quantitatively.[27] Qualitative assessments of heterogeneity involve evaluating the studies for design-related and result-related heterogeneity.[30] The assessment for design heterogeneity looks at 4 aspects of the original studies: (1) the subjects, (2) the interventions, (3) the outcomes, and (4) the methods. It is generally not appropriate to quantitatively combine data from studies that are very different from one another in relation to these characteristics.[27,29,30]

When qualitatively assessing the results of the primary studies for heterogeneity, the reviewer examines (1) the degree of overlap of the confidence intervals around the studies' point estimates of effect and (2) the disparity between the point estimates themselves. If the results of the study differ greatly, it may not make sense to combine them.[2] If, however, the reviewer decides to combine them, there are statistical approaches designed to deal with heterogeneity. Hatala and colleagues[30] state that studies should not be combined quantitatively if there are highly disparate (very different) point estimates with very little overlap of their confidence intervals. There are also statistical approaches that can be used to test the heterogeneity of the results of the studies being considered for quantitative synthesis. If it is deemed inappropriate to combine the results statistically (quantitatively), the findings should be described qualitatively.

The quantitative synthesis of outcomes from studies in the review is meta-analysis. All or only a portion of the included studies may be combined to provide summary results for the review.[27] The advantage of meta-analysis is that combining the samples from the individual studies increases the overall sample size and, therefore, the power to detect treatment effects, as well as the precision of the effects.[8] The large sample size can also allow the reviewer to examine treatment effects in subgroups of subjects.

The *Cochrane Handbook* recommends that meta-analysis follow basic principles.[29] It should typically occur as a 2-stage process. First, a summary statistic such as effect size or relative risk (RR) is calculated for each study to be combined. Then, a summary effect estimate is calculated. Studies may be weighted during this process; the larger the weight, the more a study contributes to the average effect. The 2 statistical models used to combine data in a meta-analysis are the fixed effects model and the random effects model.[26] The fixed effects model is based on the assumption that the true effect of an intervention is the same (in both direction and magnitude) in all the studies included in the meta-analysis (ie, fixed across studies). This assumption implies that observed differences among study results are due solely to chance and that there is no statistical heterogeneity.[29] If heterogeneous results (statistically heterogeneous) are included in the review, a random effects model should be used for the meta-analysis. A random effects meta-analysis model is based on an assumption that the effects being estimated in the different studies are not identical but follow some distribution. This model assumes that the studies included in the review are a random sample from a theoretical universe of all studies and that their results are randomly placed around some central value.[27] In addition to the summary effect statistic, the standard error of the pooled effect can be used to calculate a confidence interval. The confidence interval provides an estimate of the precision (certainty versus uncertainty) of the summary estimate.[29] Wider confidence intervals indicate greater uncertainty (less precision) in the estimated effect.

The summary effect measure calculated during the meta-analysis will generally be determined by how the outcomes were reported in the primary studies being combined.[31] When the study outcomes were dichotomous (binary; eg, the presence or absence of the outcome of interest), the 3 summary statistics most often reported are the odds ratio (OR), RR, and risk difference (RD). The OR is the ratio of the odds (likelihood of the outcome) in the treatment group to the odds for the control group. RR is the ratio of the risk of the outcome in the treatment to the risk in the control group. RD, also sometimes called the absolute risk reduction, is the difference between the observed risks (proportions of individuals with the outcome of interest) in the treatment and control groups.[8,29,31] When the pooled studies' outcomes are continuous, the summary statistics most often reported are the mean difference (weighted mean difference [WMD]) and standardized mean difference (SMD). The WMD reports the intervention effect as the weighted difference in the mean values of the intervention and control groups.[27] When the means from the individual studies are combined during the meta-analysis, the mean of each study is most often weighted by the precision of its estimate of effect (ie, the inverse of its variance) but may also be weighted by the sample size.[32] Weighted mean difference should be used only when all of the pooled studies used the same scale to measure the outcome of interest.[29] The SMD (also called the effect size) expresses the size of the treatment effect in each trial relative to its variability. It does not require that all studies measure the outcome in the same way.[29]

$$\text{SMD} = \frac{\text{Difference in mean outcome between groups}^{29}}{\text{Standard deviation of outcome among participants}}$$

Graphics are often used to help present the results of a meta-analysis. The typical graph, a Forest plot, displays the effect estimates and confidence intervals for each pooled study as well as the meta-analysis.[29] The results from the individual studies are generally shown as squares representing

WJ350303_258-265.qxp 4/22/08 2:28 AM Page 263

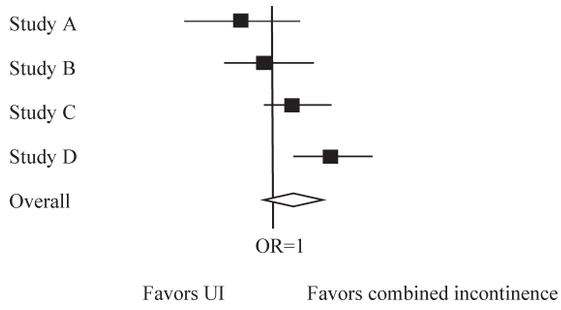

FIGURE 1. Forest plot of findings of a fictitious meta-analysis examining the likelihood of developing a pressure ulcer among nursing home residents with combined fecal and urinary incontinences versus those with only urinary incontinence (none of the estimates presented are from actual studies).

the point estimate (eg, OR of effect size) for each study. The corresponding confidence interval (usually 95%) is shown as a horizontal line running through the square. The overall estimate from the meta-analysis and its confidence interval are often represented as a diamond (the method of representation varies) at the bottom of the plot. The pooled point estimate (eg, pooled OR) is the center of the diamond, and its confidence interval is represented by the horizontal tips of the diamond.[33] The vertical line dividing the plot represents no effect (an OR or RR of 1 or an RD, WMD or SMD of 0).

Point estimates (individual and pooled) to the left of the line (your right when looking at the page) generally indicate that the experimental intervention was more effective than the control (comparison) intervention (or that the outcome was more likely to occur in the group with the characteristic of interest than in those without it), while those to the right (your left) generally indicate that it was less effective (or that the outcome was less likely to occur in the group with the characteristic of interest than in those without it). The reviewer should, however, label the plot so this is clear. If the horizontal (confidence interval) lines cross the vertical line of the plot, it generally indicates that the difference in the treatment and control groups was not statistically significant at a $P = .05$ level (95% confidence interval). The plot allows the reader to view the data that went into the meta-analysis as well as the results of the pooled analysis. The same procedures are used to analyze and present data on subgroups if there are also questions/aims addressing the effect(s) in subgroups of subjects.

Figure 1 shows a Forest plot from a fictitious review where the outcomes were dichotomous and reported as ORs. The clinical question in this systematic review with meta-analysis was, "Are nursing home residents with combined (fecal and urinary) incontinence more likely to develop pressure ulcers than those who have only urinary incontinence (UI)?" Four fictitious studies were part of this review. Each examined characteristics of nursing home patients who had pressure ulcers, including combined incontinence and UI alone. In 2 of the studies (A and B), the point estimate (OR) is to the left of the vertical line (when looking at the plot), suggesting that the subjects in these fictitious studies who had combined incontinence were less likely to develop pressure ulcers than those with only UI. The confidence intervals of both studies, however, cross the vertical line, which represents an OR of 1, indicating no statistically significant difference in the likelihood of developing a pressure ulcer in the 2 groups (those with combined incontinence and those who had only UI). In studies C and D, the point estimates are located to the right of the vertical line (when looking at the page), suggesting that subjects with combined incontinence were more likely to develop pressure ulcers than those with only UI. The confidence interval in studies C, however, also crosses the vertical line, which should be interpreted as an OR that is not statistically significantly different from 1. In contrast, in study D, the line representing the confidence does not cross the vertical line. In this study, subjects with combined incontinence were significantly more likely to have pressure ulcers than those with only UI. When the results of these studies were pooled during the meta-analysis, the point estimated (combined OR) was to the right of the vertical line, but the confidence interval crossed the vertical line. If data from this fictitious meta-analysis were to be used to answer the clinical question posed, the answer would be "no"; nursing home residents with combined incontinence are not significantly more likely to develop pressure ulcers than those who had only UI.

Drawing Conclusions Based on the Data

In the final step in the review, like in all research studies, the reviewer discusses the findings of the review and draws conclusions based on them. If the findings of the review are not consistent with the findings of studies included in the review or with previous reviews, possible reasons for these differences should be discussed. There should also be a discussion of the limitations and potential biases of the review and the implications of the review both for clinical practice and future research.

Reporting Systematic Reviews and Meta-Analyses

In an attempt to improve the quality of published meta-analyses, the Quality of Reporting of Meta-analyses (QUOROM) conference was convened to address the issues related to reporting this type of study. The focus of the conference was on meta-analyses of RCTs, but many of the recommendations are also applicable to systematic reviews without meta-analysis and to reviews that

WJ350303_258-265.qxp 4/22/08 2:28 AM Page 264

TABLE 3.

Summary of QUOROM Recommendations for Meta-Analyses of RCTs

Title identifying the report as a meta-analysis or systematic review

Structured **abstract** that includes the purpose of the review and a brief summary of the methods (data sources, selection criteria, method of validity assessment, data abstraction, study characteristics, and method of data synthesis), results, and conclusions

Body of the paper

Introduction section including a description of the clinical problem, the biological rationale for the intervention, and the rationale for the review

Methods section that provides detailed information on the search strategies used, the criteria and method used to select studies, the criteria and process used to assess the quality of the studies included, the process used to abstract data from the studies, the types of studies (designs, sample characteristics, interventions, and outcomes) included, how clinical heterogeneity was addressed and how data were quantitatively synthesized (including how missing data were handled and statistical heterogeneity was assessed)

Results including a flowchart showing the flow for the review from the identification of potentially relevant studies to the studies that contributed to the pooled effect estimate, a description of each of the studies that were part of the meta-analysis, and results of the quantitative data synthesis

Discussion summarizing the key findings, clinical inferences in light of the limitations, interpretation of the findings in light of current evidence, potential biases, and implications for future research.

Abbreviations: QUOROM, Quality of Reporting of Meta-Analyses; RCT, randomized controlled trials.

TABLE 4.

Appraisal of Critical Appraisal: Questions to Conside

- Was there a clearly focused clinical (review) question (purpose)?
- Was the search strategy comprehensive enough that all clinically relevant studies were likely to be located?
- What bibliographic databases were used?
- Were additional strategies utilized to locate potentially relevant studies?
- What years were searched?
- What languages were searched?
- Was gray literature was included in the search?
- How were studies selected for review?
- Were inclusion criteria reported?
- Were they likely to result in clinically relevant articles being identified?
- Did at least 2 individuals review articles for selection?
- What process did they use to reach agreement on article selection?
- Was the validity (quality) of studies assessed?
- Were validity criteria reported?
- How were the reviews conducted?
- Was the likelihood of bias minimized?
- Did at least 2 individuals review articles?
- What process did they use to reach agreement about the quality of the studies and data extracted?
- Were studies similar enough to combine?
- Were reasons for any differences (methodological or statistical heterogeneity) explored?
- How was heterogeneity handled?
- Were data analyzed appropriately?
- Were results clearly reported?
- If a summary measure was given, how large was the overall effect (eg, treatment effect or measure of association)?
- Are the results clinically meaningful?
- Based on the confidence intervals of the summary measure (if provided), how precise were the results?
- Was a summary of findings provided?
- If this systematic review is being used to guide clinical care:
- Were all clinically important outcomes considered?
- Were the subjects included in the review similar to the patients to whom you provide care?

WJ350303_258-265.qxp 4/22/08 2:28 AM Page 265

include other types of studies. Table 3 provides a summary of the QUOROM recommendations of what should be included in the report of a meta-analysis of RCTs.

Before deciding whether or not to use the findings of a systematic review to guide clinical care, clinicians need to critically appraise the quality of the review and the applicability of its findings to their patient population. In doing this, the clinician should examine the extent to which the review was clearly reported. Other questions to consider during the appraisal are listed in Table 4.

Conclusion

High-quality systematic reviews can be powerful tools to support clinical decision making, as well as summarize current knowledge in relation to an area of research interest. Like all research studies, systematic reviews need to be based on a structured and valid methodology and measures should be taken to minimize bias.

References

1. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Br J Surg*. 2000;87(11):1448–1454.
2. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ*. 1997;315(7121):1533–1537.
3. Phillips B, Ball C, Sackett D, et al. Oxford Centre for Evidence-Based Medicine Levels of Evidence. http://www.cebm.net/index.aspx?o=1025. Accessed January 17, 2008.
4. Physicians AAFP. Levels of evidence in AFP. http://www.aafp.org/online/en/home/publications/journals/afp/afplevels.html. Accessed 2008.
5. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs*. 2003;12(1):77–84.
6. Fleisher LA, Bass EB, McKeown P. American College of Chest Physicians guidelines for the prevention and management of postoperative atrial fibrillation after cardiac surgery. *Chest*. 2005;128:17S–23S.
7. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. Evidence-Based Medicine Working Group [see comment]. *JAMA*. 2000;284(10):1290–1296.
8. Akobeng AK. Understanding systematic reviews and meta-analysis. *Arch Dis Child*. 2005;90(8):845–848.
9. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions [see comment]. *Ann Intern Med*. 1997;126(5):376–380.
10. Bent S, Shojania KG, Saint S. The use of systematic reviews and meta-analysis in infection control and hospital epidemiology. *Am J Infect Control*. 2004;32:246–254.
11. Schmidt LM, Gotzsche PG. Of mites and men: reference bias in narrative review articles: a systematic review. *J Fam Pract*. 2005;54(4):334–338.
12. Ravnskov U. Quotation bias in reviews of the diet-heart idea. *J Clin Epidemiol*. 1995;48(5):713–719.
13. Gotzsche PC. Reference bias in reports of drug trials. *Br Med J Clin Res*. 1987;295:654–656.
14. Klassen TP, Jadad AR, Moher D. Guides for reading and interpreting systematic reviews: I. Getting started [see comment]. *Arch Pediatr Adolesc Med*. 1998;152(7):700–704.
15. Magarey JM. Elements of a systematic review. *Int J Nurs Pract*. 2001;7(6):376–382.
16. Mullen PD, Ramirez G. The promise and pitfalls of systematic reviews. *Annu Rev Public Health*. 2006;27:81–102.
17. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med*. 1997;127:380–387.
18. Garcia J, Subirana M, Sola I, Urrutia G, Bonfill X. Stomatherapists versus staff nurses for stoma care (protocol). The Cochrane Library 2007:1–5 http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/rel0001/CD004698/pdf_fs.html. Accessed January 17, 2008.
19. Meade MO, Richardson WS. Selecting and appraising studies for a systematic review. *Ann Intern Med*. 1997;127(7):531–537.
20. Conn VS, Valentine JC, Cooper HM, Rantz MJ. Grey literature in meta-analyses. *Nurs Res*. 2003;52(4):256–261.
21. Dickersin K, Min YI. NIH clinical trials and publication bias. *J Curr Clin Trials* 1993;Document no. 50:[4967 words; 4953 paragraphs] http://ottawagroup.ohri.ca/docs/NIH_clin_trials_pub_bias.pdf. Accessed January 17, 2008.
22. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research [see comment]. *Lancet*. 1991;337(8746):867–872.
23. Gerber S, Tallon D, Trelle S, Schneider M, Juni P, Egger M. Bibliographic study showed improving methodology of meta-analyses published in leading journals 1993–2002. *J Clin Epidemiol*. 2007;60:773–780.
24. McAuley L, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta–analyses? *Lancet*. 2000;356:1228–1231.
25. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. The Cochrane Library. 2007;4 http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/MR000010/frame.html. Accessed January 17, 2008.
26. van Tulder M, Furlan A, Bombardier C, Bouter L. Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine*. 2003;28(12):1290–1299.
27. Moher D, Jadad AR, Klassen TP. Guides for reading and interpreting systematic reviews: III. How did the authors synthesize the data and make their conclusions? *Arch Pediatr Adolesc Med*. 1998;152(9):915–920.
28. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17(1):1–12.
29. The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.6; 2006 http://www.cochrane.org/resources/handbook/Handbook4.2.6Sept006.pdf. Accessed January 17, 2008.
30. Hatala R, Keitz S, Wyer P, Guyatt G, for the Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine: 4. Assessing heterogeneity of primary studies in systematic reviews and whether to combine their results. *Can Med Assoc J*. 2005;172(5):661–665.
31. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med*. 1999;18(3):321–359.
32. Group B. Weighted mean or weighted mean difference. http://www.jr2.ox.ac.uk/bandolier/booth/glossary/wmd.html. Accessed January 29, 2008.
33. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ*. 2001;322(7300):1479–1480.