

Selecting Studies for Systematic Review: Inclusion and Exclusion Criteria

Timothy Meline

The University of Texas—Pan American, Edinburg, TX

The search of multiple databases to locate every study that potentially can be used to determine the efficacy of intervention is one of the first steps in the systematic review process. The search process is based on the eligibility criteria that reviewers establish before they begin the process of identifying, locating, and retrieving the research needed to address the problem of evidence-based practice. The eligibility criteria specify which studies will be included and which will be excluded from the systematic review—though the criteria may be subject to change as the systematic review progresses through the early stages of the process, some of the criteria are fundamental to collecting a rigorous and defensible set of data for the review. The criteria used for including and excluding studies form the *operational definition* of the problem (Abrami, Cohen, & d’Apollonia, 1988), and they provide a clear guideline as to

the standards of research that will be used to determine the efficacy of speech and language interventions.

The eligibility criteria are liberally applied in the beginning to ensure that relevant studies are included and no study is excluded without thorough evaluation. At the outset, studies are only excluded if they clearly meet one or more of the exclusion criteria. For example, if the focus of review is children, then studies with adult participants and no children are summarily excluded because they are outside the group of interest. Otherwise, studies are included in the pool for detailed examination at a later time. At this point, reviewers might ask which studies in the pool are relevant to the purpose of the intervention under review. This question may be the most important one that reviewers attempt to answer (cf. Gliner, Morgan, & Harmon, 2003). As you will see later (Schwartz & Wilson, 2006), the problem of identifying, locating, and retrieving this pool of studies is no small task.

Early forms of systematic reviews first appeared nearly 30 years ago in the form of meta-analyses and served as a solution to the problem of integrating the research on a specific topic (Glass, 2000). Systematic review methods have been subject to considerable discussion and debate—especially regarding the selection of studies to include or exclude from review. As Khan and Kleijnen (n.d.) advised, the choice of inclusion and exclusion criteria should logically follow from the review question and should be straightforward. However, the controversy centers on how broad or narrow the selection process should be. This is an important debate because the selection process determines the scope and validity of the systematic reviewers’ conclusions. Glass argued for the broad approach to selecting studies for review—also known as the *traditional approach*. Glass believed that “meta-analyses must deal with all studies, good bad and indifferent.”

ABSTRACT: The process of selecting studies for systematic review and meta-analysis is complex, with many layers. It is arguably the most important and perhaps the most neglected aspect in the process of integrating research on a specific topic. It is also a contentious process with opposing schools of thought as regards critical issues surrounding study selection. The debate that centers on the selection process is important because the inclusion/exclusion of studies determines the scope and validity of systematic review results. The development of inclusion/exclusion criteria is discussed, and steps in the study selection process are followed from initial evaluation to the final acceptance of studies for systematic review.

KEY WORDS: study selection criteria, study quality grading, study evaluation, inclusion/exclusion criteria, systematic review

An alternative to the traditional approach was articulated in Slavin's (1987) best-evidence principle that proposed to include only those studies that meet some high methodological standard of quality—also known as the *critical evaluation approach*. The critical evaluation approach aims to include studies that meet a predetermined threshold of quality, and it excludes those studies that do not. Lam and Kennedy (2005) explained the importance of critical evaluation as follows:

The results of a meta-analysis are only as good as the quality of the studies that are included. Therefore, the critical step in a meta-analysis is to formulate the inclusion criteria for selecting studies. If the inclusion criteria are too broad, poor quality studies may be included, lowering the confidence in the final result. If the criteria are too strict, the results are based on fewer studies and may not be generalizable. (p. 171)

Although both traditional and critical evaluation approaches have merit, adherence to either approach may impose serious limitations on meta-analyses. Selection criteria that are too narrow may severely limit the clinical application of results—an *over-exclusion threat*. On the other hand, selection criteria that are too broad may make the comparison and synthesis of studies difficult if not impossible by combining markedly different studies and introducing bias from poorly designed studies—an *over-inclusion threat*. As an alternative, some systematic reviewers advocate an intermediate approach to selecting studies for systematic review—an approach that considers the merits of both positions (cf. Abrami et al., 1988).

DEVELOPING INCLUSION AND EXCLUSION CRITERIA

Andrews, Guitar, and Howie's (1980) summary and meta-analysis represented an early attempt to integrate research on a specific topic in communication disorders. They sought to integrate the effects of treatment on stuttering as reported in the 1964–1980 time period. Their summary was criticized for being too narrow in its approach to selecting studies (Ingham & Bothe, 2002), though others argued that the selection criteria were justified (Howell & Thomas, 2002).

Andrews et al. (1980) attempted to answer the question of how effective stuttering treatment is. Their selection criteria included studies with a clinical focus and pretest/posttest research designs but excluded studies with less than 3 participants. They also excluded studies that failed to report sufficient sample statistics or sufficient raw data for calculating effect size—the common metric for combining study outcomes. Although Andrews et al. identified 100 studies that potentially met their broad criteria for eligibility, only 29 studies met all of their inclusion criteria and none of their exclusion criteria. Most if not all of the excluded studies failed to report sufficient data for calculating effect sizes. However, their result appears to be consistent with systematic reviews in other areas. According to Chambers (2004), systematic reviewers often exclude a large proportion of studies—sometimes 90% or more. Studies are typically excluded from the pool of studies

because they (a) clearly meet one or more of the exclusion criteria, (b) include incomplete or ambiguous methods, (c) fail to meet a predetermined threshold for quality, or (d) fail to report sufficient statistics or data for estimating effect sizes.

Prospective studies for systematic reviews are evaluated for eligibility on the basis of *relevance* and *acceptability* (Robey & Dalebout, 1998). Systematic reviewers ask: Is the study relevant to the review's purpose? Is the study acceptable for review? Systematic reviewers then formulate inclusion and exclusion criteria to answer these questions. Each systematic review has its own purpose and questions, so its inclusion and exclusion criteria are unique. However, inclusion and exclusion criteria typically belong to one or more of the following categories: (a) study population, (b) nature of the intervention, (c) outcome variables, (d) time period, (e) cultural and linguistic range, and (f) methodological quality.

Study Population

A systematic review requires that explicit descriptions of its methods and procedures meet a standard of *transparency* for the reader. That is, the descriptions have to be clear and precise enough that anyone could replicate the review and obtain the same studies, calculate the same treatment effects, and theoretically come to similar conclusions. To satisfy this requirement, the pertinent characteristics of the study population are described in detail. This is especially important for clinicians who ask if their client would have been eligible for this study. If the answer is *no*, the results may not be applicable for those clients' needs. Pertinent characteristics of the study population may include features such as adults or children, gender, grade level, clinical diagnosis, language, geographic region, or disability. Gender is an especially relevant characteristic of the study population when children who stutter are participants because according to Curlee and Yairi (1998), more girls than boys recover from stuttering. Thus, gender is a potential *moderator variable*—a categorical variable other than the treatment variable that explains a significant amount of the variance between studies in a systematic review.

An example of eligibility criteria is the National Institute of Neurological Disorders and Stroke's (n.d.) notice recruiting participants for a clinical trial titled *Study of Brain Activity During Speech Production and Speech Perception*. The inclusion criteria specified for the experimental group were (a) right-handed children and adolescents, (b) native speakers of American English, and (c) stuttering or phonological processing disorders. The comparison (control) group consisted of normally developing right-handed children and adolescents who were native speakers of American English. Exclusion criteria were (a) language use in the home other than American English, (b) speech reception thresholds greater than 25 dB, and (c) contraindications to magnetic resonance scanning. In a similar fashion, systematic reviewers specify inclusion and exclusion criteria for synthesizing studies, but the criteria are usually much broader.

A limitation that systematic reviewers sometimes face is a shortage of relevant studies—those that address the purpose of the review. For example, there are few studies reporting the treatment of childhood stuttering with matched or randomly assigned untreated control groups (Curlee & Yairi, 1998). Thus, if the review purpose is to assess the effects of interventions for childhood stuttering, the inclusion criteria might need to be expanded to include a variety of research types such as quasi-experimental designs.

Another limitation when attempting to integrate intervention studies for children who stutter or who are disfluent is the *definitional problem* (Ingham & Cordes, 1998) of just what constitutes a stuttering moment. The operational definitions for what constitutes stuttering and normal disfluency vary widely. Table 1 provides examples of operational definitions for stuttering that have been selected from studies with children as participants. Some of the definitions in Table 1 are quantitative and others are qualitative. In all, they illustrate the difficulty of establishing an operational definition across studies that is both useful for a systematic review and functional for interpreting the results of the included studies.

Ultimately, systematic reviewers—especially those who study treatment efficacy—value external validity as highly as internal validity (Slavin, 1987). Thus, systematic reviewers who choose an intermediate approach strive to include as many studies as possible without jeopardizing internal validity. In regard to external validity, systematic reviewers ask how representative the study sample is relative to the population of all possible studies. In regard to internal validity, systematic reviewers ask if the study's population is clinically similar enough to justify statistically combining the results (Laupacis, 2002).

Nature of the Intervention

Nature of the intervention is particularly important if the reviewer addresses the question of treatment efficacy. In this case, reviewers may ask if the studies are sufficiently similar clinically in terms of the nature of the intervention. To answer this question, systematic reviewers report the relevant features of the interventions of interest—which may include (a) operational definitions for interventions;

(b) length, timing, and intensity (dosage) of interventions; and (c) examples of interventions that are included and those that are excluded.

Outcome Variables

Systematic reviews that address questions about fluency are likely to find a variety of outcome measures represented in the study population—both quantitative and qualitative ones. Trautman, Healey, and Norris (2001) reported percentage of stuttering-type behaviors as their outcome measure. Hancock et al. (1998) chose percentage of syllables stuttered as their outcome measure. The outcome measures included in the Andrews et al. (1980) review were stuttering frequency, judgments of severity, measures of speech rate, self-reports of stuttering severity, questionnaires of attitude and speech-related behaviors, and other subjective measures. Although the outcome measure is not typically a criterion for inclusion, it is important that the outcomes be clearly presented so that a determination can be made early in the inclusion process as to the appropriateness of the study for the area under review.

Time Period

Systematic reviewers ask what the relevant time period within which studies will be selected is. For example, if the review question is limited to contemporary studies, reviewers may choose a time frame such as the prior 10 years. However, a narrow time frame may severely limit the number of eligible studies. Alternatively, the time frame may be selected on the basis of a point in time when a particular controversy emerged or a new intervention was introduced. Whatever time period is selected, reviewers are expected to provide sufficient justification for their choice.

Cultural and Linguistic Range

According to Lipsey and Wilson (2001), meta-analyses often exclude studies that are reported in languages other than English simply because of the practical difficulty of translation. Systematic reviewers may ask what the *cultural and linguistic range* of studies to be included in the review is. Cultural and linguistic range is usually reflected in the

Table 1. Operational definitions for stuttering selected from studies with children as participants.

<i>Study</i>	<i>Operational definition of stuttering</i>
Au-Yeung, Howell, & Pilgrim (1998)	Diagnosed by speech-language pathologists
Güven & Sar (2003)	Within-word disfluencies ≥ 5 per 150 words
Hancock et al. (1998)	Unnatural hesitation, interjections, restarted or incomplete phrases, and unfinished or broken words
Ryan (2000)	More than 3 stuttered words per minute
Trautman et al. (2001)	State guidelines for fluency disorders (not specified)

language and place of publication. Thus, studies that are published in the United States are usually restricted to American culture and language. Excluding non-English studies limits the scope and validity of results and may introduce *publication bias* (Khan & Kleijnen, n.d.). Publication bias is a threat to content validity if relevant studies—such as studies reported in a language other than English—are systematically excluded from the review. In any case, if reviewers choose to restrict the cultural and linguistic range of a review, they should justify the decision in relation to the purpose of the systematic review.

Methodological Quality

Methodological quality depends on (a) the type of research design and (b) the manner in which the research study is conducted. In regard to type of research design, randomized controlled trials (RCTs) are inherently the strongest design for answering questions of causality. Thus, to answer questions about the effect of intervention on disfluencies, RCTs are accepted as the *gold standard*. However, although RCTs are strong in terms of internal validity, they are often weak in terms of external validity because participants may not be representative of the broader clinical population. For example, women, elderly, and members of minority ethnic groups are often excluded from clinical trials (Gliner et al., 2003; Laupacis, 2002). Whether or not other types of research designs are included in the review is a decision the reviewer needs to make before embarking on the review. There are other issues related to analysis and interpretation when a variety of research designs are included in the systematic review, but the choice of design inclusion criteria is fundamental to the purpose question for the review.

In regard to the manner in which research is conducted, RCTs are not all conducted with the same care and precision. For example, RCTs may differ in their implementation of randomization, blinding, attrition, and allocation concealment (cf. Moher et al., 1998). In any case, there may be few if any RCTs available for the reviewer to answer questions regarding clinical efficacy—such as questions regarding stuttering treatments (Curlee & Yairi, 1998). Thus, as a matter of practicality, systematic reviewers are likely to include studies with different designs and methodologies. For this reason, Chambers (2004) recommended that reviewers code studies according to their research types. Coding studies by research type and other important variables permits statistical analysis to test for differences and evaluate the data for potential impact on the intervention effect.

Inasmuch as all research types—experimental, quasi-experimental, and others—vary in terms of methodological quality, systematic reviewers may choose to assess the quality of individual studies and code them accordingly. Although some systematic reviewers—mostly traditionalists—dismiss quality assessment procedures as unreliable, Greenwald and Russell (1991) concluded that “investigators can be in relative agreement as to the severity and seriousness of a threat to the design quality of a study. Such threats can be reliably coded, individually, and in terms of

an index of global methodological quality” (p. 23). Systematic reviewers can use quality assessment a priori as eligibility criteria to select the study pool, or they may use quality assessment to weight studies for ex post facto analysis. The point to make here is that research design is a critical element of the inclusion decision and must be clearly defined at the outset of the review.

Assessing the quality of studies. A common obstacle to assessing the quality of studies is *methodological reporting*. Methodological reporting is sometimes incomplete or ambiguous—making assessment difficult or impossible. Some potentially relevant studies may have to be discarded because they fail to report important details such as the steps taken to avoid threats to internal validity. If sufficient information about methodology is available, reviewers can assess the quality of studies by using quality indicators (Jadad, Moore, Carroll, Jenkinson, Reynolds, Gavaghan, & McQuay, 1996; Moher, Jadad, Nichol, Penman, Tugwell, & Walsh, 1995; Moher et al., 1998; Rosenthal, 1991). Quality assessment instruments typically include one of the following: (a) individual aspects of study methodology such as blinding and randomization, (b) quality checklists, or (c) quality scales that provide quantitative estimates of overall study quality (Khan, ter Riet, Popay, Nixon, & Kleijnen, n.d.). For example, Jadad et al. (1996) developed an instrument with the following 11 items:

1. Was the study described as randomized?
 2. Was the study described as double blind?
 3. Was there a description of withdrawals and drop-outs?
 4. Were the objectives of the study defined?
 5. Were the outcome measures defined clearly?
 6. Was there a clear description of the inclusion and exclusion criteria?
 7. Was the sample size justified (e.g., power calculation)?
 8. Was there a clear description of the interventions?
 9. Was there at least one control (comparison) group?
 10. Was the method used to assess adverse effects described?
 11. Were the methods of statistical analysis described?
- (p. 7)

There are two general approaches to assessing the quality of studies: the threshold approach and the quality-weighting approach. The *threshold approach* is the less inclusive of the two approaches. For example, the Agency for Healthcare Research and Quality (AHRQ, 2002) synthesized studies on speech and language evaluation instruments. They included studies based on the threshold approach. The AHRQ (2002) operational definitions for the inclusion and exclusion of studies were as follows:

Acceptable: research or analyses were well conducted, had representative samples of reasonable size, and met our psychometric evaluation criteria [reliability and validity] discussed earlier.

Unacceptable: studies were poorly conducted, used small or nonrepresentative samples, or had results that did not meet or only partially met the psychometric criteria. (p. 3)

In principle, the threshold approach guarantees a minimum level of quality (Khan et al., n.d.). To ensure an explicit description of the procedure, Khan et al. recommended: “The weakest study design that may be included in the review should be clearly stated in the inclusion/exclusion criteria in the protocol” (p. 4). A problem with implementing the threshold approach is that the decision to include or exclude studies is not always straightforward. To alleviate this problem, Abrami et al. (1988) placed studies along a continuum of confidence from *obviously meets to obviously fails to meet* the eligibility criteria, and they proceeded to include studies that reasonably met the inclusion criteria. However, this approach could bias the results in the direction of the review conclusions—an *inclusion error* (Egger & Davey Smith, 1998; Egger, Davey Smith, & Schneider, 2001).

The *quality-weighting approach* is a more inclusive approach that avoids the possibility of selection biases. It provides the benefit of a large pool of studies, fuller representation of the available research on a topic, and an opportunity to empirically examine the relationship between methodology and study outcomes (Lipsey & Wilson, 2001; Moher et al., 1998). Although selection bias is minimized, bias in assigning quality weights is a potential threat. The quality-weighting approach assesses each study and assigns a weight based on a preselected instrument. For example, quality weights might be assigned to individual studies based on an ordinal scale from 1 (lowest quality) to 5 (highest quality). With quality weights in hand, systematic reviewers are able to evaluate the data for the presence or absence of a moderator variable. Systematic reviewers ask if study quality is a variable that explains a significant amount of the variance between studies in the systematic review.

THE STUDY SELECTION PROCESS

Step 1: Apply Inclusion/Exclusion Criteria to Titles and Abstracts

The search process generates a *bibliography of candidate studies* that typically includes titles and abstracts of potentially relevant studies. At the outset, the integrity of the study selection process is evaluated by (a) piloting the inclusion/exclusion criteria on a subset of studies from the bibliography of candidate studies, and (b) testing the reliability of evaluators’ decisions. Piloting the inclusion/exclusion criteria is done to ensure that studies can be classified correctly. As a result of piloting, the inclusion/exclusion criteria may be modified to better identify relevant studies. The inclusion/exclusion criteria are subject to change throughout the selection process, but as changes are made, they must be applied retroactively to all citations in the bibliography of candidate studies.

To establish reliability in the decision-making process, two or more evaluators independently apply the inclusion/

exclusion criteria to a subset of studies from the bibliography of candidate studies. Based on the results, points of disagreement are examined. Systematic reviewers expect a high degree of reliability in the decision-making process. In this regard, Khan and Kleijnen (n.d.) observed:

Many disagreements may be simple oversights, whilst others may be matters of interpretation. These disagreements should be discussed, and where possible resolved by consensus after referring to the protocol. If disagreement is due to lack of information, the authors may have to be contacted for clarification. Any disagreements and their resolution should be recorded. (p. 4)

Step 2: Eliminate Studies That Clearly Meet One or More Exclusion Criteria

At this stage of the selection process, the emphasis is on excluding studies that clearly meet the exclusion criteria. Studies are eliminated from the bibliography of candidate studies if the titles and abstracts clearly disqualify them. The abstracts found in journal databases typically include (a) a statement of the problem, (b) a description of participants, and (c) specification of the experimental design. However, abstracts in conference programs sometimes lack essential information. For example, the title *Immediate Subjective/Objective Effects Of A Speecheasy® Device Fitting On Stuttering* was retrieved from the *ASHA Convention Abstract Archive*. The following abstract accompanied the title:

An investigation of the immediate effects of a fitting with the SpeechEasy® device on stuttering: determining the SpeechEasy’s® effect on stuttering behaviors by comparing participant speech samples in baseline, placebo, and experimental conditions. Participant perceptions pre and post and correlation of post-perceptions with actual changes in stuttering frequency are discussed. (Bartles & Ramig, 2004)

The abstract specified the independent variable, dependent variables, and experimental design but not participants. If an abstract is inconclusive, the citation remains in the bibliography of candidate studies for further evaluation after the full text is retrieved.

Step 3: Retrieve the Full Text of the Remaining Studies

At this stage, a full text of the studies that were identified in Step two and were not excluded are retrieved. The full text of reports is necessary to ensure the accuracy of decisions to include or exclude studies from the bibliography of candidate studies. Once the full texts of the studies are available, systematic reviewers proceed to Step 4.

Step 4: Evaluate the Remaining Studies for Inclusion and Exclusion

As in Step 1, the integrity of the study selection process is evaluated by testing the reliability of evaluators’ decisions. Two or more evaluators independently apply the inclusion/exclusion criteria to a subset of studies from the

bibliography of candidate studies. Points of disagreement are identified and resolved as in Step 1. If reviewers include/exclude studies based on a minimum threshold of quality, the studies are evaluated for quality. To eliminate the possibility of bias in assessing quality, author names and affiliations may be removed from reports before they are evaluated.

Step 5: Include Studies That Meet All Inclusion Criteria and No Exclusion Criteria

At this stage of the selection process, studies are further evaluated to ensure that individual studies meet all inclusion criteria and none of the exclusion criteria. In the case of studies that report incomplete or ambiguous methods, reviewers may seek further information from the original study authors. If important information is not available, a decision to exclude may be justified. If a minimum threshold of quality was established in Step 4, studies that are above the threshold are included and studies that fall below the threshold are excluded from the bibliography of candidate studies. Following this stage of the selection process, the reviewer proceeds to further exclude studies with reasons.

Step 6: Exclude Studies From Systematic Review With Reasons

At this point, studies are further excluded from the systematic review. For example, reviewers may exclude studies that do not include sufficient statistics for computing effect sizes although the studies otherwise meet the eligibility criteria. In the case of studies that report incomplete or ambiguous results, reviewers may seek further information from the authors before excluding the studies. Systematic reviewers should provide descriptions of the excluded studies along with the reasons for excluding them.

Step 7: Accept Studies for Systematic Review

In the final stage of the selection process, reviewers accept the remaining studies as eligible for systematic review. These studies constitute the sample of studies for analysis and are presumed to be representative of the population of relevant studies. The selection process ends at this point, and coding and analysis of data begin.

SUMMARY

The concept of inclusion and exclusion of data in a systematic review provides a basis on which the reviewer draws valid and reliable conclusions regarding the effect of intervention for the disorder under consideration. Not all research is created equal, and the use of evidence to guide the clinical practice needs to be cognizant of the nature and importance of the supporting research that supports various interventions. Clinicians need to understand the basis of

evidence-based decisions even if they are not engaged in the collection and analysis of the intervention studies that might guide clinical practice. Understanding what constitutes the quality characteristics of a study that is included/excluded in a summary of intervention effects is an important step in improving the quality of the clinical practice in communication disorders.

REFERENCES

- Abrami, P. C., Cohen, P. A., & d'Apollonia, S.** (1988). Implementation problems in meta-analysis. *Review of Educational Research, 58*, 151–179.
- Agency for Healthcare Research and Quality.** (2002). Criteria for determining disability in speech-language disorders. Retrieved September 10, 2005, from <http://www.ahrq.gov/clinic/epcsums/spdissum.htm>
- Andrews, G., Guitar, B., & Howie, P.** (1980). Meta-analysis of the effects of stuttering treatment. *Journal of Speech and Hearing Disorders, 45*, 287–307.
- Au-Yeung, J., Howell, P., & Pilgrim, L.** (1998). Phonological words and stuttering on function words. *Journal of Speech, Language, and Hearing Research, 41*, 1019–1030.
- Bartles, S., & Ramig, P. R.** (2004, November). *Immediate subjective/objective effects of a Speecheasy® device fitting on stuttering*. Poster session presented at the annual meeting of the American Speech-Language-Hearing Association, Philadelphia, PA.
- Chambers, E. A.** (2004). An introduction to meta-analysis with articles from *The Journal of Educational Research* (1992–2002). *The Journal of Educational Research, 98*, 35–44.
- Curlee, R., & Yairi, E.** (1998). Treatment of early childhood stuttering: Advances and research needs. *American Journal of Speech-Language Pathology, 7*, 20–26.
- Egger, M., & Davey Smith, G.** (1998). Bias in location and selection of studies. *British Medical Journal, 316*, 61–66.
- Egger, M., Davey Smith, G., & Schneider, M.** (2001). Systematic reviews of observational studies. In M. Egger, G. Davey Smith, & D. G. Altman (Eds.), *Systematic reviews in health care* (pp. 211–227). London: BMJ.
- Glass, G. V.** (2000). *Meta-analysis at 25*. Retrieved July 31, 2005, from <http://glass.ed.asu.edu/gene/papers/meta25.html>
- Gliner, J. A., Morgan, G. A., & Harmon, R. J.** (2003). Meta-analysis: Formulation and interpretation. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 1376–1379.
- Greenwald, S., & Russell, R. L.** (1991). Assessing rationales for inclusiveness in meta-analytic samples. *Psychotherapy Research, 1*, 17–24.
- Güven, A. G., & Sar, F. B.** (2003). Do the mothers of stutters use different communication styles than the mothers of fluent children? *International Journal of Psychosocial Rehabilitation, 8*, 25–36.
- Hancock, K., Craig, A., McCreedy, C., McCaul, A., Costello, D., Campbell, K., & Gilmore, G.** (1998). Two- to six-year controlled-trial stuttering outcomes for children and adolescents. *Journal of Speech, Language, and Hearing Research, 41*, 1242–1252.
- Howell, P., & Thomas, C.** (2002). Meta-analysis and scientific standards in efficacy research: A reply to Ingham and Bothe and

- Storch [Letter to the editor]. *Journal of Fluency Disorders*, 27, 177–184.
- Ingham, R. J., & Bothe, A. K.** (2002). Thomas and Howell (2001): Yet another “exercise in mega-silliness”? [Letter to the editor]. *Journal of Fluency Disorders*, 27, 169–174.
- Ingham, R. J., & Cordes, A. K.** (1998). Treatment decisions for young children who stutter: Further concerns and complexities. *American Journal of Speech-Language Pathology*, 7, 10–19.
- Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, J. M., Gavaghan, D. J., & McQuay, H. J.** (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, 17, 1–12.
- Khan, K. S., & Kleijnen, J.** (n.d.). *Selection of studies*. Retrieved August 30, 2005, from http://www.york.ac.uk/inst/crd/pdf/crd4_ph4.pdf
- Khan, K. S., ter Riet, G., Popay, J., Nixon, J., & Kleijnen, J.** (n.d.). *Study quality assessment*. Retrieved August 30, 2005, from http://www.york.ac.uk/inst/crd/pdf/crd4_ph5.pdf
- Lam, R. W., & Kennedy, S. H.** (2005). Using metaanalysis to evaluate evidence: Practical tips and traps. *Canadian Journal of Psychiatry*, 50, 167–174.
- Laupacis, A.** (2002). The Cochrane Collaboration—How is it progressing? *Statistics in Medicine*, 21, 2815–2822.
- Lipsey, M. W., & Wilson, D. B.** (2001). *Practical meta-analysis*. London: Sage.
- Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S.** (1995). Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials*, 16, 62–73.
- Moher, D., Pham, B., Jones, A., Cook, D. J., Jadad, A., Moher, M., et al.** (1998). Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analysis? *Lancet*, 352, 609–613.
- National Institute of Neurological Disorders and Stroke.** (n.d.). *Study of brain activity during speech production and speech perception*. Retrieved July 20, 2005, from <http://www.clinicaltrials.gov/show/NCT00004991>
- Robey, R. R., & Dalebout, S. D.** (1998). A tutorial on conducting meta-analyses of clinical outcome research. *Journal of Speech, Language, and Hearing Research*, 41, 1227–1241.
- Rosenthal, R.** (1991). Quality-weighting of studies in meta-analytic research. *Psychotherapy Research*, 1, 25–28.
- Ryan, B. P.** (2000). Speaking rate, conversational speech acts, interruption, and linguistic complexity of 20 pre-school stuttering and non-stuttering children and their mothers. *Clinical Linguistics & Phonetics*, 14, 25–51.
- Schwartz, J. B., & Wilson, S. J.** (2006). The art (and science) of building an evidence portfolio. *Contemporary Issues in Communication Science and Disorders*, 33 37–41.
- Slavin, R. E.** (1987). Best-evidence synthesis: Why less is more. *Educational Researcher*, 16, 15–16.
- Trautman, L. S., Healey, E. C., & Norris, J. A.** (2001). The effects of contextualization on fluency in three groups of children. *Journal of Speech, Language, and Hearing Research*, 44, 564–576.

Contact author: Timothy Meline, PhD Professor, 28 Park Place #704, Covington, TX 70433. E-mail: TM2776@aol.com