Effectiveness of Behavioral Stuttering Treatment: A Systematic Review and Meta-Analysis

Carl Herder
Courtney Howard
Chad Nye
Martine Vanryckeghem
University of Central Florida, Orlando

S ince the 1950s, researchers have investigated various stuttering treatments for both children and adults. One of the earliest studies of stuttering treatment described the use of carbon dioxide (Arthurs, Cappon, Douglass & Quarrington, 1954; Smith, 1953). Researchers found that carbon dioxide was no more effective than a placebo in improving speech fluency. Other studies published in the 1950s and 1960s explored the

ABSTRACT: The objective of this study was to conduct a systematic review and meta-analysis of the research relating to behavioral stuttering treatment. Detailed descriptions of the information retrieval, inclusion criteria, study coding, and effect size computations are provided. A total of 375 citations were identified for potential inclusion using electronic and hand-search strategies. Of the 12 included studies, six reported outcomes for treated versus nontreated participants, yielding a significant effect size of 0.91. The remaining six studies compared the effectiveness of two different treatments and yielded a nonsignificant effect size of 0.21. These data support the claim that intervention for stuttering results in an overall positive effect. Additionally, the data show that no one treatment approach for stuttering demonstrates significantly greater effects over another treatment approach.

KEY WORDS: stuttering treatment, fluency disorders, meta-analysis, evidence-based practice, systematic reviews, effect size

effects of pharmacological agents on stuttering. Some studies investigated stimulants such as methedrine, trifluoperazine, and D-amphetamine. These stimulants produced dramatic effects in reducing stuttering for a short period of time but failed to have a lasting effect (Cerciello, 1957; Fish & Bowling, 1965; Tuttle, 1952).

Other researchers examined the effects of pharmacological agents such as sedatives (Hogewind, 1940; Yannatos, 1960), tranquilizers (Holliday, 1959; Kent & Williams, 1959; Maxwell & Paterson, 1958), and neuroleptics (Andrews & Dozsa, 1977; Prins, Mandelkorn, & Cerf, 1980; Rantala & Petri-Larmi, 1976; Rosenberger, Wheelden, & Kalotkin, 1976; Swift, Swift, & Arellano, 1975; Wells & Malcolm, 1971) on the speech of people who stutter (PWS). In general, they found that the use of pharmacological treatments for stuttering produced mixed results in terms of their effect on speech improvement. In addition, the drugs typically produced negative physical and physiological side effects.

Although pharmacological interventions continue to be investigated by researchers and clinicians, they are not the focus of this article. Other types of interventions aimed at improving the speech of PWS have also been well documented in the literature. These interventions are behaviorally rather than pharmacologically based. In this article, the growing body of literature is reviewed chronologically with a view toward understanding the empirical evidence on the effects of behaviorally based interventions for PWS.

LITERATURE REVIEW

By the 1970s, there was a great deal of interest in the effectiveness of different behavioral interventions for stuttering. In a seminal review, Ingham and Andrews (1973) compiled the research on the major behavioral treatment described in the literature and thus provided a comprehensive summary and evaluation of the state-of-the-art at that time. The reviewers concluded that, among the treatments described, rhythmic speech, prolonged speech, and conditioning showed the most clinical promise. However, Ingham and Andrews expressed disappointment with the fact that many researchers and clinicians neglected to incorporate a systematic evaluation of outcomes with adequate controls for factors other than the treatment, which could affect the outcomes of PWS.

In a follow-up review, Ingham and Lewis (1978) summarized the available behavioral research and advancements in the field. The focus of the review addressed primary topics and concerns faced by clinicians, practitioners, and researchers. Despite a continued movement toward behavioral therapy as a viable intervention for PWS, Ingham and Lewis noted that a glaring lack of congruency existed relative to the basic definition of behavioral intervention and its implementation in practice. The investigators listed three components that, from their perspective, typified a behavioral approach to therapy: measures of relevant dimensions of speech behavior, therapy design, and the application of procedures that are known to modify stuttering behavior.

In the ensuing 5 years, there was a considerable shift in the volume and variety of behavioral research that examined treatments for stuttering. In addition to the existing treatments that were available in 1973, three new procedures gained notoriety and were added to the 1978 review (Ingham & Lewis): reciprocal inhibition, biofeedback, and combined behavioral and nonbehavioral procedures. Conversely, two of the treatment strategies that were reported in 1973 were excluded in the 1978 literature review: negative practice and anxiety reduction. Despite the variety of treatments available, Ingham and Lewis pointed to a discrepancy between the stated aims of the treatments under study and their actual delivery in clinical practice. These authors noted the increasing number of maintenance programs being included in treatment efficacy research, but pointed out a lack of significant follow-up data up to that point.

During the 1970s, there was a substantial growth in the body of literature on the effectiveness of stuttering intervention. Critics of the research on stuttering intervention clearly articulated their concern for the lack of rigorous data and adequate control groups to support the efficacy of behavioral treatment programs for stuttering (Ingham & Andrews, 1973; Ingham & Lewis, 1978).

The 1980s brought a significant shift in the traditional approach to investigating the effects of treatments for stuttering. Andrews, Guitar, and Howie (1980) were the first to conduct a meta-analysis of treatment studies for PWS. Andrews et al. identified 100 studies from which 29, representing 42 treatment groups, met the following criteria: (a) Authors provided a verifiable description of a treatment; (b) at least 3 PWS participated in the treatment; (c) the participants were tested pre and post treatment; and (d) sufficient data were reported to calculate effect sizes. The included studies were classified, coded, assessed, and analyzed according to 11 treatment strategies: airflow, attitude therapy, biofeedback, desensitization, gentle onset, prolonged speech, reinforcement, rhythm, self control, shadowing, and slowed speech. Results of the meta-analysis revealed that the interventions with the greatest effect were prolonged speech, gentle onset, and rhythm; and, the best predictor of treatment efficacy was the number of hours spent in therapy.

A few years later, Adams (1984) published a review of stuttering theory, research, and therapy from 1977 through 1982. He noted that by 1977, direct therapy had become an accepted intervention for children who stutter (CWS). Adams also recognized the growing clinical impact that prolonged speech had on stuttering. At that time, however, there were very few studies with large sample sizes that had investigated this practice. Furthermore, several studies provided evidence of possible short-term effects of the therapy but did not report convincing evidence of long-term effects (Andrews & Ingham, 1972; Guitar, 1976; Howie, Tanner, & Andrews, 1981; Perkins, Rudas, Johnson, Michael, & Curlee, 1974).

A 1993 issue of the *Journal of Fluency Disorders* reported the proceedings from the National Institute on Deafness and other Communication Disorders workshop on treatment efficacy research in stuttering. Mościcki (1993) mapped the parameters for designing controlled clinical trials in stuttering research. She argued that there were many critical attributes of a clinical trial, including representative sampling of participants, clearly defined selection criteria and outcomes, and long-term follow-up. She also reported that the use of randomized controlled clinical trials was the most reliable and valid way to test treatment efficacy.

Conture and Guitar (1993) described some important variables to consider in evaluating treatment efficacy when performing high-quality research with CWS. These variables included age of onset, sex ratio, ethnic and social-emotional status, and concomitant behaviors. Until this time, data on many of these variables had not been collected and, hence, empirically assessed in treatment studies. Despite their omission from the body of research, for most of the 1990s, these variables were not incorporated in many clinical trials on stuttering intervention. In addition, Curlee and Yairi (1997) discussed the importance of including control groups in efficacy research. From their perspective, most studies at that time did not meet the criteria for scientifically valid research because of the lack of nontreated controls.

Craig (1998) critically examined the research on relapse following treatment for stuttering. She argued that investigations of stuttering interventions usually resulted in impressive reductions in stuttering initially, but that these reductions were difficult to maintain. She supported the notion of high rates of relapse in PWS with a retrospective "file-drawer" study. Her review reported a significant

decrease in stuttering immediately following therapy and an increase in stuttering across groups at 12–18 months follow-up.

Thomas and Howell (2001) reviewed stuttering treatment research from 1993 to 2001 that included eight treatment studies. The authors indicated that their review contained a meta-analysis. Although this review provided useful information relative to the types of treatment studies available, the review was not a systematic one (Ingham, 2002). Furthermore, the authors compared multiple treatment studies using a vote-counting approach. Contrary to Thomas and Howell's report, the review did not combine the outcomes and their effect sizes in a meta-analysis.

In 2003, several authors collaborated to reevaluate the status of treatment efficacy research published up to that time. A common theme in the literature was the importance of differentiating between treatment outcome research and treatment efficacy research. *Treatment outcome research* was defined as research that was conducted in a "real-world" setting with functional clinical application. *Treatment efficacy research* was defined as being conducted in an ideal laboratory setting where effectiveness, efficacy, and effects were measured more systematically (Ingham, 2003; Langevin & Kully, 2003). Ratner (2005), however, expressed concern that using too narrow an interpretation of evidence might suggest that clinical decisions must reflect both empirical evidence and theory.

Summary of Review Findings

The need for and use of a scientific inquiry of behavioral stuttering treatment efficacy is well documented in the literature. In the field of communication sciences and disorders, Andrews et al. (1980) conducted the first meta-analysis study that statistically synthesized treatment effects for stuttering interventions. No doubt that this meta-analysis advanced the understanding of what works among available treatments. However, this advancement lacked a number of methodological procedures that are now considered fundamental to a systematic review and meta-analysis (Boruch, Soydan & de Moya, 2004).

The limitations of previous reviews demonstrate a need to reexamine the issue of treatment effectiveness for stuttering intervention with state-of-the-art methods for systematic reviews and meta-analysis (Cooper, 1998; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001). As such, the purpose of the present study was to conduct a systematic review and meta-analysis of the effectiveness of behavioral stuttering treatment for PWS. The procedures used have been detailed in other articles in this journal issue, and the reader is referred to them for more information.

METHOD

Information Retrieval

To locate as many relevant studies as possible, the authors conducted hand and electronic searches. Hand searches were conducted for the following relevant journals in communication sciences and disorders (years in parentheses):

- *Journal of Fluency Disorders* (1976–2004)
- *Journal of Speech and Hearing Research* (1958–1996)
- *Journal of Speech and Hearing Disorders* (1936–1990)
- *Journal of Speech, Language, and Hearing Research* (1997–2004)
- *Language, Speech, and Hearing Services in Schools* (1970–2004)
- *American Journal of Speech-Language Pathology* (1992–2004)
- *Behavior Therapy* (1970–2004)
- *Journal of Behavioral Therapy and Experimental Psychiatry* (1970–2004)
- *Journal of Communication Disorders* (1967–2004)
- *Folia Phoniatrica* (1987–1992)
- *Folia Phoniatrica et Logopaedica* (1995–2003)
- *Journal of Medical Speech Language Pathology* (1993–2003)
- *Seminars in Speech and Language* (1983–2003)

For electronic searches, databases most likely to index studies relevant to the review and most likely to catalog published and unpublished studies were searched. They included:

- PsycINFO
- ERIC
- MEDLINE
- CINHL
- C2-SPECTR
- Cochrane Central Register of Controlled Trials and Dissertation Abstracts

To locate relevant studies in each database, the following terms were used:

- Target population: stutt*, stam*,
- Intervention: therap*, interven*.
- Outcome: stutt*, fluen*, dysfl*, disfl*,

During the electronic searches, database thesauri were consulted when available to ensure that all appropriate synonyms were included in the search term categories. Search terms were modified to meet the requirements of individual databases. An attempt was also made to locate studies published in languages other than English.

Each citation abstract retrieved through an electronic search was reviewed independently by two readers to assess the appropriateness of the study for inclusion in the review. If the abstract indicated that the study met the inclusion criteria, or if the abstract was unclear, a complete copy of the study was obtained and evaluated for inclusion in the review based on the criteria discussed below.

When a complete copy of all potentially appropriate studies had been retrieved, an independent evaluation of each study was conducted by two readers for a final

inclusion decision. Disagreements between the two reviewers were reconciled by discussion until a consensus decision was reached. If consensus could not be reached, the full-text version of the study was submitted to a third reader for a final decision. Reviewers were not blinded at any level of the review to names of authors, institutions, or publication sources.

Inclusion Criteria for Studies

For a study to be included in our review, the following information must have been reported:

- The participants were diagnosed as PWS.
- The treatment method was behavioral.
- There were outcomes of speech behavior.
- The participants were randomly assigned to an experimental and control (or comparison) condition before the intervention.

Participant characteristics. The participants for an included study were those who had been diagnosed as PWS using labels such as stuttering or stammering. No studies with participants who had been diagnosed as clutterers were included. All participant age groups were included to cover both children and adult stutterers.

Treatment characteristics. All included studies used a behaviorally based intervention as the method to improve speech behaviors. No studies were included that focused on psychopharmacological interventions. No restriction was imposed regarding the length of the intervention, number of sessions, length of sessions, or frequency of treatment sessions.

Outcome characteristics. The outcomes of interest in the behavioral treatment for stuttering are typically measures of speech production. The most common outcomes included in this review are stuttered syllables or words per minute and percentage of stuttered syllables or words per minute.

Other measures of effectiveness in an individual study included an assessment of speech-related psychological and emotional outcomes as a result of the intervention. However, for the purpose of this review, we focused exclusively on the speech behaviors, such as those identified above, as the dependent measures of treatment effectiveness.

Design characteristics. Only randomized control trials (RCTs) were included in this review. Studies that used quasi-experimental group, pre-experimental group, or single-subject designs were excluded from the review. Studies that used qualitative approaches were also excluded. All included studies identified both experimental and control (or comparison) groups to which participants were assigned before intervention. A control group is defined as a nontreatment condition; a comparison group is defined as an alternative treatment condition.

Study Coding

Each included study was coded for participant, outcome, treatment, and design characteristics. The coding was done independently by two reviewers, each using a separate

complete copy of the retrieved study. When disagreements arose in the coding process, differences were resolved in the same manner presented earlier in the section on study inclusion. A complete coding form is provided in Appendix A.

Effect Size Computations

The calculation of intervention effects is a straightforward procedure of subtracting the mean of the control group from the mean of the experimental group and dividing by the averaged (pooled) groups' standard deviation as follows:

$$\overline{ES} = \frac{\overline{X}_1 - \overline{X}_2}{SD_{pooled}}$$

$\overline{ES}$ = effect size
$\overline{X}_1$ = mean of the treated group
$\overline{X}_2$ = mean of the comparison group (treatment 2 or control)
SD_{pooled} = pooled standard deviation.

The effect size provides an estimate of the effectiveness of an intervention and is reported using a *d*-index metric. For those studies that reported results using *t* tests, *F* tests, or *p* values, calculations were possible using standard formulas to derive the effect size. In addition, effect sizes were calculated using a correction factor for small sample sizes known as Hedges's *g* (Hedges & Olkin, 1984).

Synthesis of effect sizes. The purpose of a meta-analysis is to statistically synthesize the results of several studies. The synthesis produces a sample estimate of the magnitude of intervention effect in the population. When multiple results from multiple studies are combined without regard to sample size, studies with small samples are given undue weight in the averaging process. In order to synthesize data proportionate to the size of the sample, effect sizes were weighted using an inverse variance weighting factor. Through the weighting factor, each study's effect size estimate contributes proportionately to the average effect based on the study's sample size. The weighting factor used depends on the statistical model selected to estimate the average effect size. Standard practice is to estimate the average effect size using both a fixed effects and a random effects model and evaluate the results of both using a homogeneity analysis.

Homogeneity analysis. A homogeneity analysis assesses the degree to which differences in effect sizes from individual studies are due to sampling error and other factors. There are two approaches used in this assessment. The first approach is to test the null hypothesis under the fixed effects model. This analysis assumes that the effect sizes in the sample vary only by sampling error and is reported using the Q statistic. When the Q statistic is large enough to reject the null hypothesis, it is concluded that the effect sizes for each study vary beyond that which is expected by sampling error.

Publication bias. The effort to identify all relevant studies in the published and grey literature is always

susceptible to a bias, with published studies reporting larger effect sizes than grey literature studies (Rothstein, Sutton, & Borenstein, 2005). To assess the magnitude of this bias in the meta-analysis, a funnel plot is produced. A funnel plot is a graphic presentation of the potential missing studies in the meta-analysis that, if included, may alter the observed average effect size estimated through the meta-analysis.

Sensitivity analysis. The purpose of a sensitivity analysis is to test how sensitive the average effect size is to the inclusion of one particular study. To test the impact of each individual study on the overall effect, a one-study-removed analysis is conducted. This analysis is conducted by removing the effect size of each study while retaining the remaining studies and computing average effect size with that study removed.

RESULTS

A total of 1,798 articles were initially identified for this review. These articles were flagged based on title, citation, or abstract. Application of the inclusion criteria reduced the number of articles for which full texts were retrieved to 375. Review of the full texts revealed that 151 were narratives, observational studies, qualitative studies, or nontreatment studies. Of the remaining 224 treatment articles, 19 met the inclusion criteria. All other investigations were either nonrandomized quasi-experimental, multiple group, single group, or single case (i.e., case study) designed studies. Seven of the 19 studies were later excluded due to insufficient data, which resulted in 12 included studies. In these studies, the average participant was an 18-year-old male who was receiving 1 hr of intervention for approximately 12.5 weeks.

Meta-Analysis of Included Studies

Before the meta-analysis, the studies were split into two groups: (a) studies that compared a treated to a nontreated group (e.g., treatment vs. control) and (b) those that compared a treated group to another treated group (e.g., treatment 1 vs. treatment 2). At the most global level, the outcomes of each study were summarized using a single effect size. All analyses are presented using a forest plot graphic to represent the relative effect sizes, comparison groups, and outcome measures for each study.
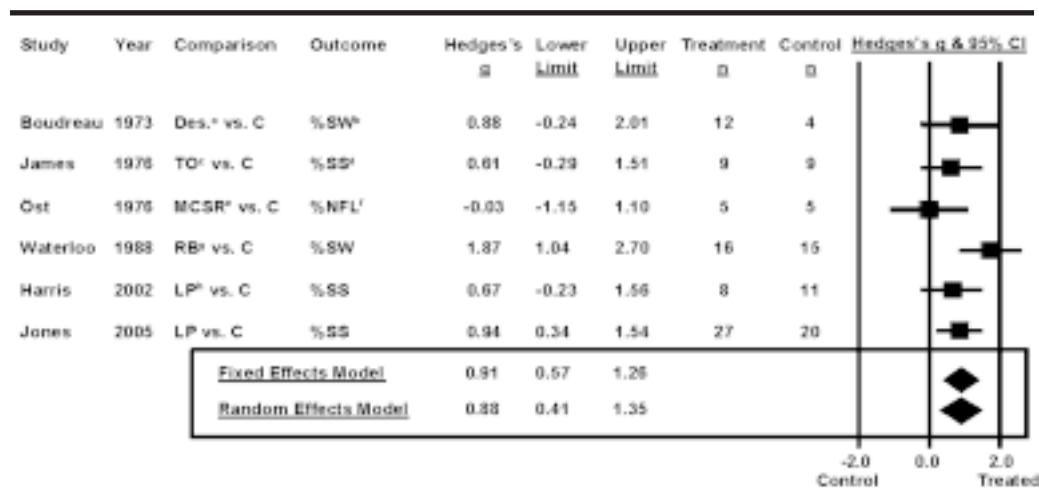
Intervention Efficacy: Treatment Versus Control

The initial analysis assessed the effects of stuttering intervention across all relevant outcomes using a fixed effects model. Figure 1 shows the overall average effect size for the treatment group compared to the control group and the studies' effect sizes from which the average effect size was derived. Figure 1 shows that the overall effect of the stuttering interventions is 0.91 *SD* in favor of the treated participants when compared to the nontreated participants. The 95% confidence interval (CI) for this effect size in the population ranges from 0.57 to 1.26 *SD*. This result is interpreted to mean that stuttering intervention is positive and statistically significant.

Homogeneity in Effect Sizes

As shown in Figure 1, four of the six studies produced a nonsignificant result (the 95% CI includes zero) while the aggregate effect indicated that intervention was statistically significant overall (the 95% CI does not include zero). One explanation for the variation in effect sizes may be due to

Figure 1. Overall effect for behavioral stuttering intervention: Treatment vs. control.

Study	Year	Comparison	Outcome	Hedges's g	Lower Limit	Upper Limit	Treatment n	Control n	Hedges's g & 95% CI
Boudreau	1973	Des.ᵃ vs. C	%SWᵇ	0.88	-0.24	2.01	12	4	
James	1976	TOᶜ vs. C	%SSᵈ	0.61	-0.29	1.51	9	9	
Öst	1976	MCSRᵉ vs. C	%NFLᶠ	-0.03	-1.15	1.10	5	5	
Waterloo	1988	RBᵍ vs. C	%SW	1.87	1.04	2.70	16	16	
Harris	2002	LPʰ vs. C	%SS	0.67	-0.23	1.56	8	11	
Jones	2005	LP vs. C	%SS	0.94	0.34	1.54	27	20	
Fixed Effects Model				0.91	0.57	1.26			
Random Effects Model				0.88	0.41	1.35			

Note. Heterogeneity statistics for a fixed model: $Q = 8.49$, $df = 5$, $p = .13$, $I^2 = 41.07$.

[a]Desensitization; [b]percentage of stuttered words; [c]Timeout from speaking; [d]percentage of stuttered syllables; [e]metronome-conditioned speech retraining; [f]percentage of nonfluency; [g]regulated breathing; [h]Lidcombe program.

differences in the studies' samples (i.e., sampling error). To assess the plausibility of this explanation, a test of homogeneity was conducted. Figure 1 shows that the Q value for the homogeneity test was 8.49 ($df = 5$, $p = 0.13$), indicating that there was no evidence to reject the null hypothesis that the effect sizes varied beyond what was expected from sampling error. However, given the small number of studies in the meta-analysis, this result could be due to a lack of statistical power, as shown in the Q statistic for studies with small sample sizes. A more informative measure of the amount of variation in the effect sizes beyond sampling error is the I^2 statistic. Figure 1 shows that 41.07% of the variation in the effect sizes was beyond what would have been expected from sampling error. This is considered a moderate amount of variation in the effect sizes, indicating that the average effect size is representative of the studies from which it was derived.

Publication Bias

As stated earlier, despite the best effort to identify all relevant studies on stuttering intervention, some studies may have been missed in the retrieval process. Had they been identified and included in the analysis, it may have altered the average effect size. To assess the probability of missing these studies and the impact they may have had on the average effect size, a funnel plot analysis was performed (see Figure 2). The funnel plot analysis uses a trim and fill procedure to assess the probability of missing studies (Duvall & Tweedie, 2000).

The trim and fill procedure is based on the number of studies included, size of the observed aggregated effect, and standard error for each study. The opened circles and diamond in Figure 2 indicate the observed effect size and standard error associated with the included studies. On the basis of the funnel plot, there is evidence to support the claim that the average effect of 0.91 SD is positive, substantial, and statistically significant but is also a conservative estimate of the effect of the stuttering interventions.

Sensitivity Analysis

Figure 1 also shows that four of the six studies produced a nonsignificant result (e.g., the 95% CI includes zero) even though the average effect size indicated that intervention overall was statistically significant (i.e., the 95% CI does not include zero). The effect sizes for these studies raises a question about whether one study's effect size may be the reason for the average effect size being positive and statistically significant. Using a one-study-removed analysis, six separate analyses were individually conducted to assess the residual impact on the overall effect size when each individual study was removed from the aggregation. Figure 3 presents the one-study-removed analysis for the six included studies.

As can be seen in Figure 3, the removal of any individual study had little impact on the magnitude of the average effect size or its CI. When any one study was removed from the calculation, the average effect size ranged from a low of 0.71 SD to a high of 1.01 SD. Thus, no single study influenced the average effect size to the point of nonsignificance when it was removed from the analysis. The results of the one-study-removed analysis show that the average effect size is a positive, stable, statistically significant estimate of the effect of stuttering interventions.

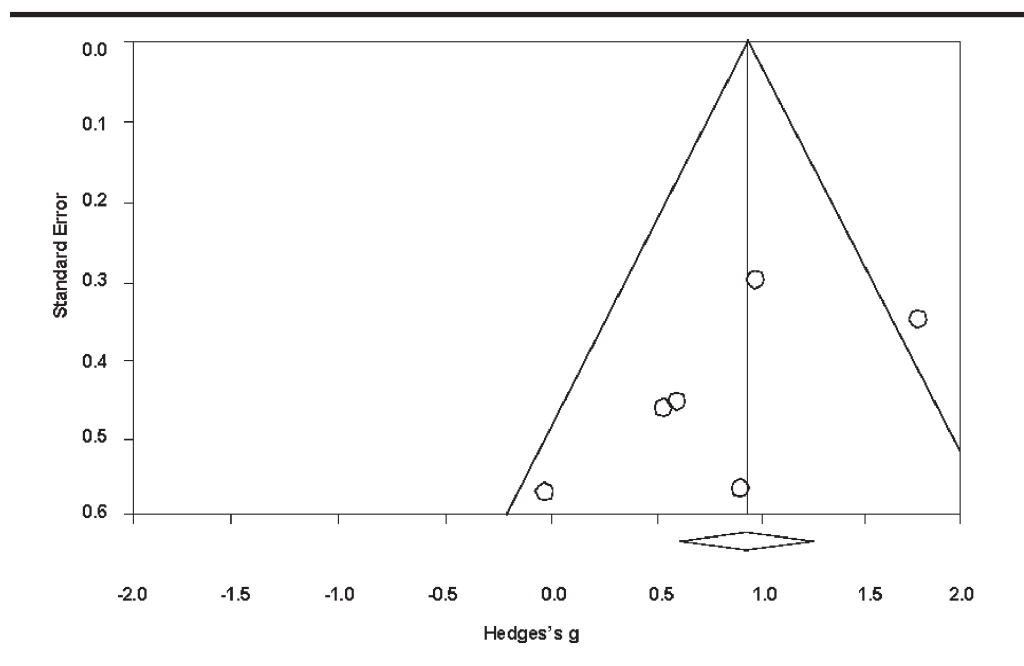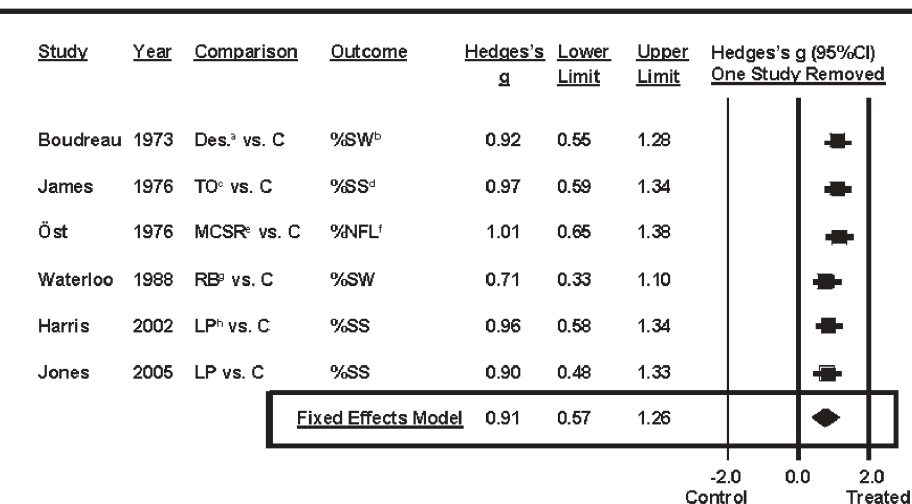**Figure 2.** Funnel plot of standard error by Hedges's *g*.

Figure 3. Overall effect for behavioral stuttering intervention: One study removed.

Study	Year	Comparison	Outcome	Hedges's g	Lower Limit	Upper Limit	Hedges's g (95%CI) One Study Removed
Boudreau	1973	Des.[a] vs. C	%SW[b]	0.92	0.55	1.28	
James	1976	TO[c] vs. C	%SS[d]	0.97	0.59	1.34	
Öst	1976	MCSR[e] vs. C	%NFL[f]	1.01	0.65	1.38	
Waterloo	1988	RB[g] vs. C	%SW	0.71	0.33	1.10	
Harris	2002	LP[h] vs. C	%SS	0.96	0.58	1.34	
Jones	2005	LP vs. C	%SS	0.90	0.48	1.33	
		Fixed Effects Model		0.91	0.57	1.26	

Overall effect of each individual study removed from the aggregate calculation

[a] Desensitization; [b] Percent Stuttered Words; [c] Time-Out from Speaking; [d] Percent Stuttered Syllables; [e] Metronome-Conditioned Speech Retraining; [f] Percent Non-Fluency; [g] Regulated-Breathing; [h] Lidcombe Program

Cumulative Effect of Stuttering Interventions Over Time

Considering the extensive history of efficacy research available in the area of stuttering, a cumulative analysis was conducted to understand at what point the additive effects of the stuttering interventions for the included studies reached a positive and significant combined effect size. In other words, was a point reached where additional studies had little impact on our understanding of the nature and degree of treatment efficacy?

The cumulative analysis starts with the initial study's effect size and averages each subsequent study's effect size, sample size, and standard error. In Figure 4, for example, Boudreau (1973) published the first RCT and reported an intervention effect of 0.88 *SD* (95% CI, –0.24 to 2.01). The second study, by James (1976), reported an effect of 0.61 *SD* (95% CI, –0.29 to 1.51). When added together, the two studies produced a combined effect size of 0.72 *SD* that was not significant (95% CI, –0.01 to 1.42). Thus, the combined outcome data of the first two studies show that behavioral interventions for stuttering were ineffective up to 1976. When the third study (Öst, 1976) was added to the Boudreau (1973) and James (1976) studies, the results revealed that the effect size increased to 0.51 *SD* and the CI narrowed (95% CI, –.09 to 1.10) but still indicated a nonsignificant average effect size. However, when the fourth study by Waterloo (1988) was added to the previous studies, the overall effect size increased substantially to 0.97 *SD* and was significant (95% CI, 0.49 to 1.46). With the addition of the last two studies, the effect size increased to 0.91 *SD* and remained statistically significant.

The results of the cumulative analysis show that there is evidence to support the claim that since 1988, the available RCTs on stuttering intervention show that the positive effect of stuttering interventions has increased. In addition, this increase has grown larger over time and has stabilized as a positive effect size.

Intervention Effectiveness: Treatment 1 Versus Treatment 2

An analysis of the studies that compared one treated group against a second treated group revealed a nonsignificant average effect size of 0.21 (CI 95%, –0.14 to 0.56). These data support the claim that when individual behavioral stuttering treatments are compared against one another, no one treatment demonstrates a superior effect over another (see Figure 5).

Even though the omnibus effect for Treatment 1 versus Treatment 2 was not statistically significant, it was possible that the age of the participants could be a moderating variable. Thus, a subgroup analysis was conducted comparing intervention effects for studies treating adolescents and adults (*n* = 4) and studies treating children (*n* = 3). Figure 6 presents the subgroup analysis indicating a nonstatistically significant improvement (*p* > .05) for each group when treatment approaches were compared.

SUMMARY AND CONCLUSION

Overall Effect of Stuttering Interventions

The treatment of PWS has been a subject of discussion for many years in the fields of communication disorders, psychiatry, psychology, and medicine. In fact, many of the

Herder et al.: *Effectiveness of Behavioral Stuttering Treatment* **67**

Figure 4. Cumulative efficacy of behavioral stuttering intervention.

Study	Year	Comparison	Outcome	Hedges's g	Lower Limit	Upper Limit	Treatment n	Control n	Hedges's g & 95% CI
Boudreau	1973	Des.[a] vs. C	%SW[b]	0.88	-0.24	2.01	12	4	
James	1976	TO[c] vs. C	%SS[d]	0.72	-0.01	1.42	21	13	
Öst	1976	MCSR[e] vs. C	%NFL[f]	0.51	-0.09	1.10	26	18	
Waterloo	1988	RB[g] vs. C	%SW	0.97	0.49	1.46	42	33	
Harris	2002	LP[h] vs. C	%SS	0.90	0.48	1.33	50	44	
Jones	2005	LP vs. C	%SS	0.91	0.57	1.26	77	64	
		Fixed Effects Model		0.91	0.57	1.26			

-2.0   0.0   2.0
Control   Treated

[a] Desensitization; [b] Percent Stuttered Words; [c] Time-Out from Speaking; [d] Percent Stuttered Syllables; [e] Metronome-Conditioned Speech Retraining; [f] Percent Non-Fluency; [g] Regulated-Breathing; [h] Lidcombe Program

founders and early scholars in the field were individuals who stuttered. This study has attempted to identify and summarize the best available research in the area of stuttering intervention by synthesizing and analyzing as many RCTs as possible using systematic review and meta-analytic techniques.

Several major findings were observed:

- PWS show large and statistically significant intervention effects when compared to those not receiving intervention.

- The individual study effects are consistent across time.
- No single treatment approach produces statistically significant superior results over another.
- Stuttering intervention research is noteworthy for the relatively small sample sizes used.

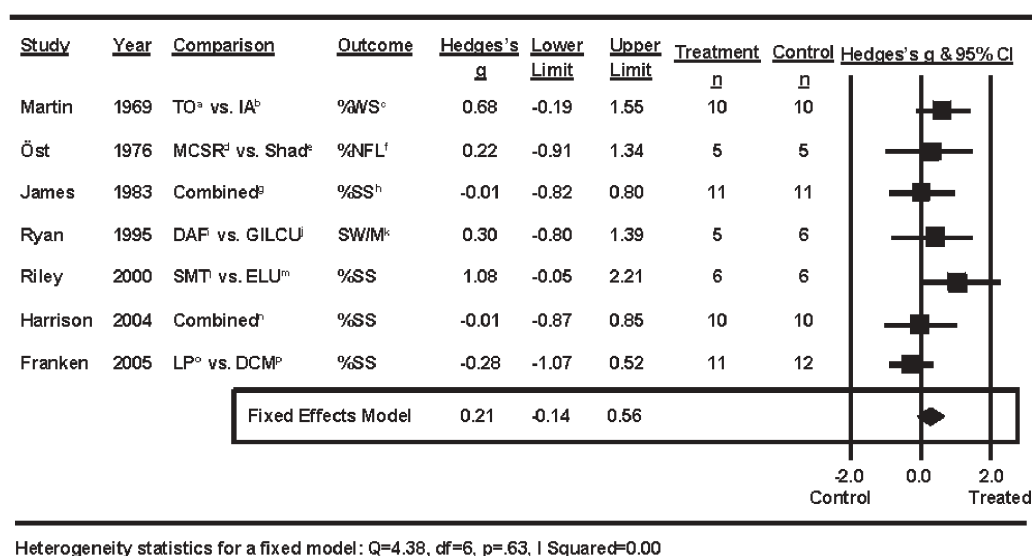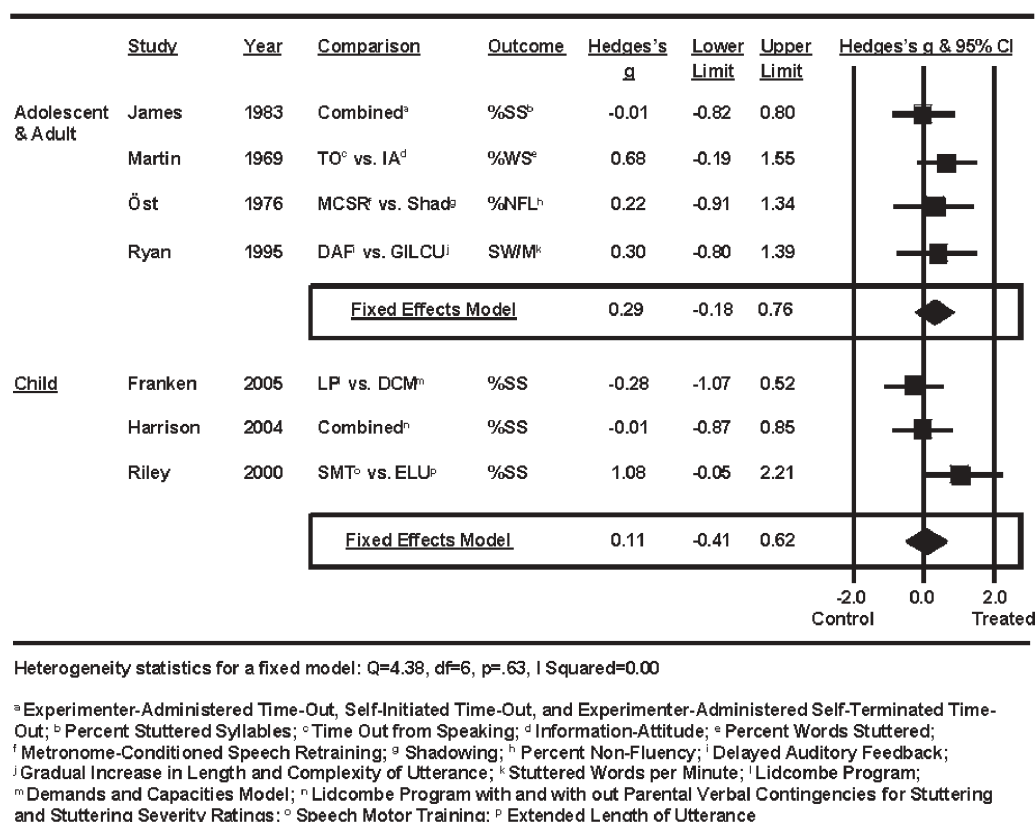**Figure 5.** Overall effect for behavioral stuttering intervention: Treatment 1 vs. Treatment 2.

Study	Year	Comparison	Outcome	Hedges's g	Lower Limit	Upper Limit	Treatment n	Control n	Hedges's g & 95% CI
Martin	1969	TO[a] vs. IA[b]	%WS[c]	0.68	-0.19	1.55	10	10	
Öst	1976	MCSR[d] vs. Shad[e]	%NFL[f]	0.22	-0.91	1.34	5	5	
James	1983	Combined[g]	%SS[h]	-0.01	-0.82	0.80	11	11	
Ryan	1995	DAF[i] vs. GILCU[j]	SW/M[k]	0.30	-0.80	1.39	5	6	
Riley	2000	SMT[l] vs. ELU[m]	%SS	1.08	-0.05	2.21	6	6	
Harrison	2004	Combined[n]	%SS	-0.01	-0.87	0.85	10	10	
Franken	2005	LP[o] vs. DCM[p]	%SS	-0.28	-1.07	0.52	11	12	
		Fixed Effects Model		0.21	-0.14	0.56			

-2.0   0.0   2.0
Control   Treated

Heterogeneity statistics for a fixed model: Q=4.38, df=6, p=.63, I Squared=0.00

[a] Time Out from Speaking; [b] Information-Attitude; [c] Percent Words Stuttered; [d] Metronome-Conditioned Speech Retraining; [e] Shadowing; [f] Percent Non-Fluency; [g] Experimenter-Administered Time-Out, Self-Initiated Time-Out, and Experimenter-Administered Self-Terminated Time-Out; [h] Percent Stuttered Syllables; [i] Delayed Auditory Feedback; [j] Gradual Increase in Length and Complexity of Utterance; [k] Stuttered Words per Minute; [l] Speech Motor Training; [m] Extended Length of Utterance; [n] Lidcombe Program with and with out Parental Verbal Contingencies for Stuttering and Stuttering Severity Ratings; [o] Lidcombe Program; [p] Demands and Capacities Model

Figure 6. Overall effect for behavioral stuttering intervention: Treatment 1 vs. treatment 2.

	Study	Year	Comparison	Outcome	Hedges's g	Lower Limit	Upper Limit	Hedges's g & 95% CI
Adolescent & Adult	James	1983	Combined[a]	%SS[b]	-0.01	-0.82	0.80	
	Martin	1969	TO[c] vs. IA[d]	%WS[e]	0.68	-0.19	1.55	
	Öst	1976	MCSR[f] vs. Shad[g]	%NFL[h]	0.22	-0.91	1.34	
	Ryan	1995	DAF[i] vs. GILCU[j]	SW/M[k]	0.30	-0.80	1.39	
	Fixed Effects Model				0.29	-0.18	0.76	
Child	Franken	2005	LP vs. DCM[m]	%SS	-0.28	-1.07	0.52	
	Harrison	2004	Combined[n]	%SS	-0.01	-0.87	0.85	
	Riley	2000	SMT[o] vs. ELU[p]	%SS	1.08	-0.05	2.21	
	Fixed Effects Model				0.11	-0.41	0.62	

Heterogeneity statistics for a fixed model: Q=4.38, df=6, p=.63, I Squared=0.00

[a] Experimenter-Administered Time-Out, Self-Initiated Time-Out, and Experimenter-Administered Self-Terminated Time-Out; [b] Percent Stuttered Syllables; [c] Time Out from Speaking; [d] Information-Attitude; [e] Percent Words Stuttered; [f] Metronome-Conditioned Speech Retraining; [g] Shadowing; [h] Percent Non-Fluency; [i] Delayed Auditory Feedback; [j] Gradual Increase in Length and Complexity of Utterance; [k] Stuttered Words per Minute; [l] Lidcombe Program; [m] Demands and Capacities Model; [n] Lidcombe Program with and with out Parental Verbal Contingencies for Stuttering and Stuttering Severity Ratings; [o] Speech Motor Training; [p] Extended Length of Utterance

Magnitude and Consistency of Intervention Effect

The 0.91 effect size that was observed for the combined intervention effect suggests that the average treated person who stutters would make a substantial improvement as a result of participation in a treatment program. Furthermore, this effect has been observed in studies over a 33-year period of reported research. Although the individual study effects are not always statistically significant as measured by the 95% CI, the fact remains that the average effect across 33 years and six RCTs is relatively constant. What is particularly interesting is that with time, the error associated with the 95% CI has decreased until, with the most recent studies, the error terms are nearly identical.

One study produced a very large and statistically significant effect size of 1.87 (Waterloo, 1988). The magnitude of this result is substantially larger than the nearest positive result, which calls for a need to replicate the study to verify the outcome magnitude. Furthermore, exploration of this particular study could provide insight into the components of the intervention that resulted in such a large effect.

The importance of the findings from this synthesis and review are substantial in that the data suggest that the clinical application of intervention techniques is becoming increasingly refined and defined. This is demonstrated by the reduction in the width of the CI for each study over time. A possible implication of the shrinking CI is that

researchers are becoming increasingly more homogeneous in their selection criteria for participants. This does not mean that the within-sample characteristics are homogeneous, but rather that the participants of one study are similar to those of another study. The result is that the clinician may gain increasing confidence that the available data on methods of intervention are based on cumulative research rather than scattered and unrelated studies.

One may argue that the use of only RCT studies as the research design of choice unnecessarily restricts the source of the evidence surrounding intervention for PWS. However, the fact that the results are so substantial offers a compelling argument that RCT stuttering intervention literature can be used as a comparative standard for other analyses of intervention for PWS. Studies comparing the magnitude of effect based on the research design employed have often reported a significant bias in favor of less rigorous group designs such as quasi-experimental, cohort, or single group pre/posttest-only designs (Moher et al., 1995). Using Cohen's scale of effect size interpretation, 0.91 reflects a large effect, which is considered substantial in human behavior interventions in general (Cohen, 1988).

Comparability of Intervention Approaches

The head-to-head comparison of different treatment cohorts provided data as to the effectiveness of particular intervention approaches. The finding that there was not a significant difference between the treatment cohorts suggests the

Herder et al.: *Effectiveness of Behavioral Stuttering Treatment* **69**

possibility that the critical element(s) for successful intervention might not lie with the intervention itself. On the basis of the current data, two major conditions may have had an impact on the effect of intervention: the intervention strategy and the specific or combined characteristics of PWS. The intervention strategy can be ruled out because the data indicate that the effect size was consistent for the one-study-removed analysis. The impact of participant characteristics was ruled out because cumulative analysis revealed that the effect size of this variable was increasingly more precise over time.

A number of interesting speculations remain that might have had an impact on the efficacy of the treatments described. One factor, clinician impact, centers around the clinician delivering the treatment as an important element in the effectiveness of the intervention. It might well be that the clinician represents a helping individual who is perceived to have the knowledge and skills to bring about a change in the speech behavior of a person who stutters. In education, this is often referred to as the Pygmalion effect, or self-fulfilling prophecy (Merton, 1957). That is, individuals are motivated to change when certain levels of expectation are set even if the expectations would not be achievable under normal circumstances. In the case of the person who stutters, the clinician establishes a level of acceptable speech functioning as the goal of intervention and the person who stutters rises to that level of expectation, thereby improving his or her speech performance.

The absence of a difference between the broadly based treatment strategies also suggests the possibility that there is (are) effective common element(s) basic to the strategies compared. Moreover, it is clear that some of the approaches compared are not fully independent of each other. For example, there are elements of time out and information attitude that are similar; shadowing and metronome pacing also have similar elements.

With respect to the current meta-analysis, we also need to keep in mind that not all treatment cohorts were compared. We have no knowledge as to the effectiveness of cohorts like delayed auditory feedback and speech motor training. Consideration should also be given to the possibility that there might not be one type of intervention that proves to be effective for all PWS. It has been reported that the topography of the behaviors evidenced by a person who stutters (e.g., exhibits more in the way of sound/syllable repetitions rather than prolongations) plays a role in the utility of particular therapeutic approaches.

As we have indicated, there are no hard data that underpin the aforementioned speculations as to why our analysis showed that there was not a significant difference between the outcomes of different treatments. However, the speculations should serve as the basis for further studies that either support or reject certain of the possibilities suggested.

Underpowered Research

Of the 12 studies that were included in this systematic review and meta-analysis, only two studies were conducted with 15 or more participants in both the treated and control/comparison groups. The potential for detecting treatment effects with such a small N is reduced proportionately in these studies. The fact that four studies out of the six reported in the treatment versus control cohort individually produced nonsignificant effect sizes argues for (a) a mandate for large N trials and (b) the need to use meta-analysis to aggregate the data across smaller N studies.

The call for more research is a common theme in research literature. The results from this study would argue not for more research, but for more research with large enough sample sizes to effectively detect the impact of the interventions being studied. Although it is certainly true that the clinical researcher must always balance the internal and external validity controls, it is also true that without a theoretically representative sample, the observed results are potentially limited regardless of internal validity.

The alternative to original research using appropriately large sample sizes is meta-analysis of small N studies. A benefit of meta-analysis is the creation of a "synthetic sample" (H. M. T. Turner, personal communication, October 15, 2005) that provides a sufficiently robust representation of the larger population of PWS. In order to verify the findings of this synthetic sample, it will be necessary to compare intervention effects of original large sample size studies to the aggregated small N study meta-analysis. The low incidence of stuttering negatively impacts the possibility of obtaining large samples. In this case, meta-analysis should be used as a substitute for high-quality original research simply because appropriately large samples are too difficult to assemble.

In order for meta-analysis to be a feasible option, specific standards should be outlined for intervention studies. Mos´cicki (1993) outlined several design issues that need to be considered for tests of treatment efficacy. She noted a need for a representative sample, a large enough sample size in order to sufficiently demonstrate differences between comparison groups, and a clear operational definition of the outcome being tested. When multiple researchers abide by the same strict research design criteria, a meta-analysis of their studies can provide highly useful information. This may allow for subgroup analyses, including the effects of age, socioeconomic status, gender, length of intervention, and more. Recently, researchers of the Lidcombe program (Harris, Onslow, Packman, Harrison, & Menzies, 2002; Harrison, Onslow, & Menzies, 2004; Jones et al., 2005) provided rigorous treatment study designs. The reader is referred to these studies for a model of high-quality designed treatment efficacy research.

It has been said that "if you continue to do what you've always done, you will continue to get what you've got." (Abrami, Bernard, Cobb, Nye, Turner, & Wade, 2004). There is a need in the area of stuttering to attend to what has been done in past research, identify the research needs to advance our understanding of treatment effects, and be alert to unexpected results. The discussion among researchers relative to the efficacy of intervention for PWS and the research methodology is perhaps more active than in any other field within communication sciences and disorders. Both the clinician and researcher need to have access to the best available treatment evidence to guide them in providing the most appropriate intervention and to study the effects of the interventions they practice.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Abrami, P., Bernard, B., Cobb, B., Nye, C., Turner, H., & Wade, A. (2005, April). *Producing C2 systematic reviews to inform evidence-based policy and practice in education.* Professional training workshop presented at the American Educational Research Association, Montreal, Canada.

Adams, M. R. (1984). Stuttering theory, research, and therapy: A five-year retrospective look ahead. *Journal of Fluency Disorders, 9,* 103–113.

Andrews, G., & Dozsa, M. (1977). Haloperidol and the treatment of stuttering. *Journal of Fluency Disorders, 2,* 217–224.

Andrews, G., Guitar, B., & Howie, P. (1980). Meta-analysis of the effects of stuttering treatment. *Journal of Speech and Hearing Disorders, 45,* 287–307.

Andrews, G., & Ingham, R. J. (1972). An approach to the evaluation of stuttering therapy. *Journal of Speech and Hearing Research, 15,* 296–302.

Arthurs, R. G. S., Cappon, D., Douglass, E., & Quarrington, B. (1954). Carbon dioxide therapy with stutterers. *Diseases of the Nervous System, 15,* 123–126.

Boruch, R., Soydan, H., & de Moya, D. (2004). The Campbell Collaboration. *Brief Treatment and Crisis Intervention, 4,* 277–287.

*Boudreau, L. A., & Jeffrey, C. J. (1973). Stuttering treated by desensitization. *Journal of Behavior Therapy and Experimental Psychiatry, 4,* 209–212.

Cerciello, P. (1957). La D-amfetamine nella cura della balbuzie [D-Amphetamine in the cure of stuttering]. *Bollettino della Società Italiana di Fonetica, Foniatria e Audiologia, 7,* 14–16.

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conture, E. G., & Guitar, B. E. (1993). Evaluating efficacy of treatment of stuttering: School-age children. *Journal of Fluency Disorders, 18,* 253–287.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Craig, A. (1998). Relapse following treatment for stuttering: A critical review and correlative data. *Journal of Fluency Disorders, 23,* 1–30.

Curlee, R. F., & Yairi, E. (1997). Early intervention with early childhood stuttering: A critical examination of the data. *American Journal of Speech-Language Pathology, 6,* 8–18.

Duvall, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463.

Fish, B. H., & Bowling, E. (1965). Stuttering. The effect of treatment with D-amphetamine and a tranquilizing agent, trifluoperazine. A preliminary report on an uncontrolled study. *Calif. Med., 103,* 337–339.

*Franken, M. C. J., Kielstra-Van der Schalk, C. J., & Boelens, H. (2005). Experimental treatment of stuttering: A preliminary study. *Journal of Fluency Disorders, 30,* 189–199.

Guitar, B. E. (1976). Pretreatment factors associated with the outcome of stuttering therapy. *Journal of Speech and Hearing Research, 19,* 590–600.

*Harris, V., Onslow, M., Packman, A., Harrison, E., & Menzies, R. (2002). An experimental investigation of the impact of the Lidcombe program on early stuttering. *Journal of Fluency Disorders, 27,* 203–214.

*Harrison, E., Onslow, M., & Menzies, R. (2004). Dismantling the Lidcombe program of early stuttering intervention: Verbal contingencies for stuttering and clinical measurement. *International Journal of Communication Disorders, 39,* 257–267.

Hedges, L., & Olkin, I. (1984). *Statistical methods for meta-analysis.* New York: Academic Press.

Hogewind, F. (1940). Medical treatment of stuttering. *Journal of Speech Disorders, 5,* 203–208.

Holliday, A. R. (1959). Effect of meprobamate on stuttering. *Northwest Medicine, 59,* 837–841.

Howie, P. M., Tanner, S., & Andrews, G. (1981). Short and long term outcome in an intensive treatment program for adult stutterers. *Journal of Speech and Hearing Disorders, 46,* 104–109.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Ingham, J. C. (2003). Evidence-based treatment of stuttering: I. Definition and application. *Journal of Fluency Disorders, 28,* 197–207.

Ingham, R. J. (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., & Andrews, G. (1973). Behavior therapy and stuttering: A review. *Journal of Speech and Hearing Disorders, 38,* 405–440.

Ingham, R. J., & Lewis, J. I. (1978). Behavior therapy and stuttering: And the story grows. *Human Communication, 3,* 125–152.

*James, J. E. (1976). The influence of duration on the effects of time-out from speaking. *Journal of Speech and Hearing Research, 19,* 206–215.

*James, J. E. (1983). Parameters of the influence of self-initiated time-out from speaking on stuttering. *Journal of Communication Disorders, 16,* 123–132.

*Jones, M., Onslow, M., Packman, A., Williams, S., Ormond, T., Schwarz, I., & Gebski, V. (2005). Randomised controlled trial of the Lidcombe programme of early stuttering intervention. *British Medical Journal, 331,* 659–661.

Kent, L. R., & Williams, D. E. (1959). Use of meprobamate as an adjunct to stuttering therapy. *Journal of Speech and Hearing Disorders, 24,* 64–69.

Langevin, M., & Kully, D. (2003). Evidence-based treatment of stuttering: III. Evidence-based practice in a clinical setting. *Journal of Fluency Disorders, 28,* 219–236.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

*Martin, R. R., & Haroldson, S. K. (1969). The effects of two treatment procedures on stuttering. *Journal of Communication Disorders, 2,* 115–125.

Maxwell, R. D. H., & Paterson, J. W. (1958). Meprobamate in the treatment of stuttering. *British Medical Journal, 1,* 873–874.

Merton, R. K. (1957). *Social theory and social structure.* Glencoe, IL: Free Press.

Moher, D., Pham, B., Jones, A., Cook, D. J., Jadad, A. R., Moher, M., et al. (1998). Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses?, *Lancet, 352,* 609–613.

Mościcki, E. K. (1993). Fundamental methodological consider-ations in controlled clinical trials. *Journal of Fluency Disorders, 18,* 183–196.

*Öst, L., Götestam, K. G., & Melin, L. (1976). A controlled study of two behavioral methods in the treatment of stuttering. *Behavior Therapy, 7,* 587–592.

Perkins, W. H., Rudas, J., Johnson, L., Michael, W. B., & Curlee, R. F. (1974). Replacement of stuttering with normal speech: III. Clinical effectiveness. *Journal of Speech and Hearing Disorders, 39,* 416–428.

Prins, D., Mandelkorn, T., & Cerf, F. A. (1980). Principal and differential effects of haloperidol and placebo treatments upon speech disfluencies in stutterers. *Journal of Speech and Hearing Research, 23,* 614–629.

Rantala, S. L., & Petri-Larmi, M. (1976). Haloperidol (serenase) in the treatment of stuttering. *Folia Phoniatrica, 28,* 354–361.

Ratner, N. B. (2005). Evidence-based practice in stuttering: Some questions to consider. *Journal of Fluency Disorders, 30,* 163–188.

*Riley, G. D., & Ingham, J. C. (2000). Acoustic duration changes associated with two types of treatment for children who stutter. *Journal of Speech, Language, and Hearing Research, 43,* 965–978.

Rosenberger, P. B., Wheelden, J. A., & Kalotkin, M. (1976). The effect of haloperidol on stuttering. *American Journal of Psychiatry, 133,* 331–334.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention assessment and adjustments.* Chichester, England: Wiley.

*Ryan, B. P., & Van Kirk Ryan, B. (1995). Programmed stuttering treatment for children: Comparison of two establish-ment programs through transfer, maintenance, and follow-up. *Journal of Speech and Hearing Research, 38,* 61–75.

Smith, A. M. (1953). Treatment of stutterers with carbon dioxide. *Diseases of the Nervous System, 14,* 243–244.

Swift, W. J., Swift, E. W., & Arellano, M. (1975). Haloperidol as a treatment for adult stuttering. *Comprehensive Psychiatry, 16,* 61–67.

Thomas, C., & Howell, P. (2001). Assessing efficacy of stuttering treatments. *Journal of Fluency Disorders, 26,* 311–333.

Tuttle, E. (1952). Hyperventilation in a patient who stammered: Methedrine as an adjunct to psychotherapy. *Am. J. Med., 13,* 777–779.

*Waterloo, K. K., & Götestam, K. G. (1988). The regulated-breathing method for stuttering: An experimental evaluation. *Journal of Behavior Therapy and Experimental Psychiatry, 19,* 11–19.

Wells, P. G., & Malcolm, M. T. (1971). Controlled trial of the treatment of 36 stutterers. *British Journal of Psychiatry, 119,* 603–604.

Yannatos, G. (1960). L'hydroxyzine dans la thérapeutique des bégaiements [Hydroxyzine in the treatment of stuttering]. *Journal Français d'Oto-Laryngologie et Chirurgie Maxillo-Faciale, 9,* 293–296.

Contact author: Carl Herder, 4320 Center Key Road #2317, Winter Park, FL 32792. E-mail: cherder@gmail.com

APPENDIX A. CODING GUIDE

Publication Source

Subject Characteristics
        Group n
        Age
        Gender: percent male/female
        Socio-economic status
        Sample source:
        Setting
        Severity of participants pre-test
        Race/ethnicity

Intervention Characteristics
        Treatment setting (Individual/group)
        Primary type of treatment:
        Total length of intervention program
        Length of time per day of participation activity
        Number of sessions
        Intervention implemented as described

Outcome Measure
        Type of measure
        Who administered the measure?

Design Characteristics
        Participant recruitment procedure
        Subject assignment procedure
        Method of random assignment (pg _____)
        Blinding

Effect Size Source
        Mean & SD t test F test p value

Method of Analysis