Tutorial

Defining, Identifying, and Evaluating Clinical Trials of Stuttering Treatments: A Tutorial for Clinicians

Mark Onslow

Australian Stuttering Research Centre, The University of Sydney

Mark Jones

Queensland Clinical Trials Centre, University of Queensland, Australia

Sue O'Brian

Ross Menzies

Ann Packman

Australian Stuttering Research Centre, The University of Sydney

Contact author: Mark Onslow, Australian Stuttering Research Centre, The University of Sydney, Gate 2 East Street, Lidcombe, NSW, Australia. E-mail: m.onslow@usyd.edu.au .

Abstract

Purpose: To develop a method for clinicians to evaluate stuttering treatment efficacy research with very little burden of work.

Method: The clinical trial is the most fundamental, clinically interpretable, and useful output unit of stuttering treatment research. We define a clinical trial of a stuttering treatment and specify 3 levels of clinical trials evidence. We use this taxonomy to identify and evaluate clinical trials of stuttering treatment. Our taxonomy draws on 2 fundamental principles of clinical trials used to evaluate health care: randomization and effect size.

Results: Published clinical trials of stuttering treatments were identified and allocated to 1 of 3 levels of evidence.

Conclusions: We outline a 3-step, semi-automated, Internet-based method to identify the publication of a report of stuttering treatment efficacy. For a report identified as such, a 10-item checklist is applied to verify its status as a clinical trial and to allocate it to 1 of 3 levels of clinical trials evidence. The present taxonomy reduces the burden of work of a 136-item checklist in an existing taxonomy.

Key Words: stuttering severity, children, adults, clinical trials, efficacy

Scholars in the field of stuttering research can serve clinicians with reviews of research dealing with the efficacy of various treatments. Such reviews can assist clinicians to deliver evidence-based treatments to their clients. There have been several efforts to review stuttering treatment reports that clinicians might draw on. These reviews have used several different criteria for determining whether a report warrants clinical consideration, and several different methods to evaluate the merits of a report.

Andrews, Guitar, and Howie (1980) drew on biostatistical measurement of effect size in evaluating reports available at the time. Reports were considered if they had at least 3 participants and pretreatment and posttreatment data were available. For 42 nonpharmacological reports that met those criteria, Andrews et al. used effect size to judge the worth of treatments designed to alleviate stuttering. Another approach involves simply cataloguing the fundamental features of each treatment report, such as the number of participants and their ages, the follow-up period, summarized outcomes, and so on (Bloodstein, 1995; R. J. Ingham, 1984; St. Louis & Westbrook, 1987). This method does not focus on systematic quantification of the value of each report, presumably with the intention that clinicians who consult such catalogues would select treatments for closer inspection at their publication source. Subsequently, clinicians would make up their own minds about what treatments are of value and which they might use as a first choice.

Arguably, the most useful approach to reviewing reports of stuttering treatment efficacy includes some method to evaluate their quality. Bloodstein's (1995) well-known criteria are one example of this approach, although it was not applied in that text to the cataloguing of treatment reports. More recently, Bothe, Davidow, Bramlett, and Ingham (2006) presented a method to identify strengths and weaknesses of reports and to provide guidance for researchers. Bothe et al. developed five "methods criteria" and five "outcomes criteria." For the most part, for a report to be included, four of the five methods criteria needed to be met. This occurred in 39 of 162 reports. Subsequently, the number of outcomes

criteria met determined the methodological quality of a report. A related article implemented the Bothe et al. guidelines with a checklist-based Stuttering Treatment Research Evaluation and Assessment Tool (STREAT; Davidow, Bothe, & Bramlett, 2006). The STREAT checklist contains 136 items, ranging across seven major headings and 15 subheadings, with many items requiring a prose response. The purpose of the present article is to develop an alternative taxonomy for clinicians to evaluate stuttering efficacy research with much less burden of work.

We present the clinical trial as the most fundamental, clinically interpretable, and useful output unit of stuttering treatment research, and present a definition of a clinical trial of a stuttering treatment. We specify three levels of clinical trials evidence, and we use this taxonomy to identify and evaluate clinical trials of stuttering treatment, at the time of writing, for four age groups. Historically, recommendations for how to evaluate stuttering treatment efficacy research are derived from a discipline-and-disorder-specific focus, drawing on the consensus of various authorities in speech-language pathology writings on stuttering (e.g., Bloodstein, 1995; Bothe, Davidow, Bramlett, & Ingham, 2006; Conture & Guitar, 1993; Curlee, 1993; Davidow et al., 2006; J. C. Ingham & Riley, 1998; R. J. Ingham, 1984; Onslow & Packman, 1999; Starkweather, 1993). Our definition of a clinical trial of a stuttering treatment draws on, and is generally consistent with, that body of work. However, our taxonomy to evaluate clinical trials of stuttering treatment draws on two fundamental principles used to evaluate health care in general: randomization and effect size.

We outline a method for clinicians to be immediately alerted to the publication of a clinical trial and evaluate it with very little burden of work. We outline a partly automated, three-step, Internet-based method to identify the publication of a report of treatment efficacy. For a report identified as such, a 10-item checklist is applied to verify its status as a clinical trial and to allocate it to one of three levels of clinical trials evidence.

Because our overall purpose is to reduce the burden of work for clinicians in identifying and evaluating clinical trials of stuttering treatments, we have attempted to improve the readability of this article with the use of footnotes. These notes present, separate from our narrative, much of the background statistical detail and reasoning, and details of the logic underlying our recommendations. We begin by defining and delimiting our recurring term "efficacy."[1]

Defining Clinical Trials

An Entire Treatment

Bothe, Davidow, Bramlett, and Ingham (2006) referred to a notion of "trial quality." For clinical purposes, we present instead the notion of a speech-

language pathology clinical trial.2 In this context, we mean a clinical trial to be an evaluation of the efficacy of an entire intervention. We exclude from consideration as a clinical trial any report that does not explore the efficacy of an entire treatment. An "entire treatment" refers to treatments described in their entirety in a report or manualized independently. An "entire treatment" includes the pragmatic application of a predetermined treatment package, normally involving programmed instruction, such as in the treatment program reported by Boberg and Kully (1994). An "entire treatment" also includes reports that involve the consistent application of variables to alleviate stuttering, without using programmed instruction, with the intention of so doing until a clinically meaningful response to that intervention occurs. An example of the latter would be Martin, Kuhl, and Haroldson's (1972) nonprogrammed application to children of time-out.

We exclude from consideration reports in which portions of treatments are evaluated experimentally rather than to determine efficacy of the entire package. For example, G. D. Riley and Ingham (2000) compared extended length of utterance treatment with speech motor training to determine the acoustic effects of each treatment. However, the treatments were not presented in their entirety. Likewise, J. Riley and Riley (1999) reported on a group who received a standard 12 weeks of speech motor training that reduced their stuttering by around half, rather than determining the effects of the entire treatment. Also, the Harris, Onslow, Packman, Harrison, and Menzies (2002) and Franken, Kielstra-Van der Schalka, and Boelens (2005) reports are excluded because they presented children with a 12-week portion of two manualized treatments that normally require longer than that period to be efficacious. Likewise, the Lattermann, Euler, and Neumann (2008) report compared 16 weeks of a treatment to 16 weeks of no treatment.

We do not mean that the results of clinical experiments, and other clinical research, are of no clinical interest. To the contrary, they may be useful in a clinical repertoire of knowledge about what treatments to consider. For example, the Harris et al. (2002) report provides information that a certain treatment has an impact over a 12-week period that is clinically superior than natural recovery. Below we show how such clinical experimentation permits preliminary statistical inference to supplement clinical trials evidence in stuttering. However, such experiments do not provide the fundamental information of interest to a clinician: what the outcome might be for a client if a treatment is administered in its entirety. This brings us to the topic of outcomes.

Outcomes

In clinical trials, outcomes are the intended health benefit of a treatment. Clinical trials require the specification of primary outcome. In medical health research,

primary outcomes for potentially fatal diseases may include postmorbid lifespan, and for nonfatal diseases the primary outcomes may include a measure of quality of life. For disorders with biochemical components, such as diabetes, primary outcomes may comprise biochemical data. For clinical trials in speech-language pathology, primary outcomes are commonly, but not necessarily, measures of speech or language.

How Many Outcomes?

For the "gold standard" randomized controlled trial (RCT), it is advisable to have one primary outcome. Publication of an RCT in many journals is not allowable if this criterion is not met (The CONSORT Group, 2008). Statistical inference in an RCT is normally based on one primary outcome, and having more than one invalidates that statistical inference, rendering the trial uninterpretable. This is because the primary hypothesis of an RCT is linked to the primary outcome, which is used to set the participant numbers for analysis of trial data (for details pertinent to clinical trials in stuttering, see Jones, Gebski, Onslow, & Packman, 2002). At most, two primary outcomes are considered to be acceptable (The CONSORT Group, 2008).

This does not mean that publication of an RCT necessarily produces sparse information involving one or two outcomes. To the contrary, in RCTs it is standard practice to specify a priori a number of "secondary outcomes" that are of interest in addition to the primary ones. Those secondary outcomes may be used to provide additional, rich information about the health benefits of the treatment. Nor does the restriction on primary outcomes in RCTs preclude more than one outcome being used in less stringent trial methods, which are discussed below.3

The Temporal Dimension of Outcomes

The notion of trial outcome includes time; assessment of a change in health without a passage of time is meaningless. Hence pretreatment and follow-up data are needed in trials.4 In order to be deemed a clinical trial of a stuttering treatment, the present definition requires a report to have at least one follow-up outcome measure.

This raises the issue of the period that should be used for a follow-up. The recurring suggestion by authorities cited above of at least 6–12 months seems reasonable. However, for the present purposes, such a suggestion might not be ideal. This is because clinical trials document not only the potential value of treatments but also when treatments may have no value or limited value. Such findings are particularly important, especially if they contradict prior positive findings.

Consider scenarios where researchers show in a trial that a treatment does not alleviate stuttering in a clinically significant way (e.g., Block, Onslow, Packman, Gray, & Dacakis, 2005; Huber, O'Brian, Onslow, & Packman, 2003).5 For such a demonstration of no treatment benefit, a follow-up period is not required. If after 1 month or 2 weeks, or even on the last day of treatment, there are no apparent reductions in stuttering, then a conclusion is possible from the trial. Hence our definition of a clinical trial of stuttering treatment specifies no follow-up period in the event of no treatment effect. We specify in our definition a minimum follow-up period of 3 months in the case of a treatment effect.6

Outcomes Beyond the Clinic

We agree with authorities cited above that outcome measures beyond the clinic are essential in clinical trials of stuttering treatment. Within-clinic speech samples may be biased because of discriminated learning to the clinic (Stokes & Baer, 1977). Further, although data suggest that follow-up stuttering scores are the same within and beyond the clinic (Block et al., 2005; R. J. Ingham, 1980; Packman, Onslow, O'Brian, & Huber, 2004), it is simply not believable to claim a treatment effect from clinic data. As expressed by Davidow et al. (2006), without such data the "real world effectiveness" (p. 130) is not demonstrated. We add that real-world assessments require conversational speech.

Bias

Bias is the bane of science, and particularly so in the case of treatment efficacy research. This is because any attempt to assess stuttering behavior could be reactive to assessment. The most common effect of such bias is a belief that a treatment is efficacious when it is not, or a belief that a treatment is more efficacious than it really is. In other words, a result of bias in efficacy research is overestimation of the outcome of a treatment that misleads clinicians about its value.

Independent, Prospective Observations

We agree with Bothe, Davidow, Bramlett, and Ingham (2006) about the value of a bias-freed independent observer of stuttering during the pretreatment to follow-up period. Observations need to be independent of treatment. Further, observations need to be based on either video- or audio-recorded speech samples so blinded observation is possible; an observer in real time is likely to know whether treatment has occurred.7

Some stuttering treatment reports are retrospective in the sense that participants had received their treatments at the time the trial was planned. We exclude such reports consideration as a clinical trial. For example, Koushik, Shenker, Onslow, and Adaman (2008) and Miller and Guitar (2006) reported on clinical caseloads followed up with audio and video recordings. Such follow-up

reports of clinical caseloads are of interest, but the retrospective methodology admits possible biases in participant selection that are not present in prospective methods. Bloodstein (1995) expressed this issue by saying, "Cases written up for publication are apparently chosen for the precise reason that they were successes" (p. 401).

These requirements of independent, recording-based speech observations also exclude clinical file audit information. This is most commonly reported in a retrospective fashion by consulting clinic treatment records, such as the large cohort studies of Jones, Onslow, Harrison, and Packman (2000), Kingston, Huber, Onslow, Jones, and Packman (2003), and Webster (1980), and smaller case studies such as Yaruss, Coleman, and Hammer (2006) and Starkweather and Gottwald (1993). This is not meant to negate the contributions of retrospective file audits. The Jones et al. (2000) and Kingston et al. (2003) reports contribute information about the ideal age at which to introduce the treatment concerned, and the Yaruss et al. (2006) report is a welcome initial development of a treatment (Onslow & Yaruss, 2007). However, such reports present measures collected by clinicians during the clinical process, and this admits sources of bias. Stuttering behavior is a complicated set of events (e.g., Ambrose & Yairi, 1999; Onslow, 1995; Teesson, Packman, & Onslow, 2003), and skilled observers are needed for such judgments in research. Inexperienced clinicians may grossly underestimate the number of stuttering moments (Brundage, Bothe, Lengeling, & Evans, 2006). Additionally, during the treatment process, clinicians have detailed familiarity with clients' speech, and their judgments of the number of stuttering moments are likely to be influenced by the views of their clients or clients' parents.

Definition of a Clinical Trial of a Stuttering Treatment

In summary, our definition of a clinical trial of a stuttering treatment is as follows: A clinical trial of a stuttering treatment is (a) a prospective attempt to determine the outcome or outcomes of (b) at least one entire treatment with (c) at least one pretreatment and one follow-up outcome of at least 3 months in the case of a reported positive outcome, and (d) where outcomes involve speech observations that are independent of treatment and derived from recordings of conversational speech beyond the clinic.

Evaluating Clinical Trials of Stuttering Treatment

Evaluating a clinical trial is a critical component in the cycle of evidence-based clinical reasoning, because all clinical trials evidence is not equal. In this section we draw on the concepts of randomization and effect size to provide a simple and usable method to evaluate a clinical trial.

Randomization

Randomization is the most effective way to deal with bias in clinical trials.8 In effect, a nonrandomized clinical trial is a clinical trial with one group only. A nonrandomized trial does not compare one treatment to another treatment or to no treatment. The method is simply pretreatment and follow-up observations of one treatment. Even in the case of a nonrandomized clinical trial that involves several treatments, the design amounts to little more than a number of independent, nonrandomized trials. Probably the best example of this design in stuttering research is the Craig et al. (1996) nonrandomized trial of home-based speech restructuring, clinic-based speech restructuring, and electromyographic (EMG) biofeedback treatment.

As outlined below, one-group trials are essential in development of clinical trials evidence. However, the bias in such designs causes overestimation of the health benefits of a treatment. Placebo effects and natural recovery are potential sources of bias in nonrandomized trials. Another problem is that stuttering involves natural variations in severity, particularly in the case of preschool children. There is a tendency for those who stutter to present at clinics when their disorder is severe, with subsequent improvement in their condition more likely than not from natural severity fluctuation.9

In a nonrandomized trial, the only option presented to participants is to receive a particular treatment, so the decision of whether a participant enrolls in the trial rests completely with the participant and the clinician. This introduces what is known as allocation bias. The majority of clinical trials in stuttering have occurred in the context of clinical service provision facilities. It seems inevitable, then, that clinicians have introduced bias by enrolling participants in trials when there is a priori reason to believe that the treatment would be a useful and successful one.

Another source of bias with nonrandomized trials is the loading of case history variables that are likely to affect treatment outcome. In the case of speech-restructuring treatments for adults, such a variable would be a history of successful treatment followed by relapse. The rate at which this occurs is generally estimated to be between 30% and 50% (Boberg & Kully, 1994; Howie, Tanner, & Andrews, 1981; Martin, 1981; Perkins, 1981) and as high as 73% if relapse is defined by self-report for a clinically significant period (Craig & Hancock, 1995). Hence, participants in a trial of speech-restructuring treatment would likely have had a successful treatment previously but relapsed. Clearly, such participants are prone to show a favorable response in the trial but to relapse again after the follow-up period. In the case of stuttering preschoolers, parents are likely to bring children for treatment soon after stuttering onset. Considering that time since onset is the best known clinical predictor of natural recovery, natural recovery would be expected to be a serious bias in the results

of a nonrandomized trial of a treatment for early stuttering. As we shall see shortly, this prediction appears to be true.

Drop-outs are when participants allocated to receive a treatment in a trial either do not receive the treatment, receive part of the treatment, or are not assessed for the entire follow-up period. Drop-outs are inevitable in clinical trials and are a worrying source of bias in nonrandomized trials of stuttering treatment (e.g., Bloodstein, 1995; Bothe, Davidow, Bramlett, & Ingham, 2006; Curlee & Yairi, 1998; Davidow et al., 2006; Hegde, 2007; Thomas & Howell, 2001). Although such drop-outs may be reported, they lead to dramatic overestimates of the benefits that can be expected of a treatment. One reason drop-outs occur in trials is poor outcome, and the results of the trial simply exclude these cases. An example would be a nonrandomized trial by Onslow, Costa, Andrews, Harrison, and Packman (1996) where 6 participants dropped out because they could not learn the requisite speech pattern in the time allocated. Hence, the reported outcome obviously was an overestimate of effect size.10

As stated by Robey (2005), the randomized trial is a "tremendously complex and expensive proposition requiring extraordinary scientific rigor" (Robey, 2005, p. 6). Indeed, in order to have such a trial considered for publication in many top medical journals, a 22-item methodological checklist is a prerequisite (The CONSORT Group, 2008). In return for this rigor and cost, the randomized trial removes the sources of bias listed above that occur in nonrandomized trials (but not all sources of bias). Bias associated with placebo effects, natural recovery, and regression to the mean is eliminated with the random allocation of participants to one or more treatment groups or a control group. The randomization process eliminates allocation bias because neither participants nor clinicians decide which participant receives which treatment. Ideally, the randomization process is conducted by personnel independent of the study. Several clinical trials centers in Australia offer randomization services to those who conduct clinical trials.

Effect Size

Arguably the strongest feature of randomized trials is that they provide the best estimate of the health outcome obtained with a treatment. This may be measured with various indices of effect size (Lipsey & Wilson, 1993). One such index is the odds ratio. This figure gives the odds of someone, in the population of interest, who receives the treatment being better off than someone who does not. For example, in an RCT of the Lidcombe Program (Jones et al., 2005), the odds ratio at 9 months postrandomization was 7.7, which meant that 9 months after the start of the treatment a child who received the Lidcombe Program had 7.7 higher odds of stuttering with severity below 1.0%SS during everyday conversation than a child who did not receive the Lidcombe Program.11

In the event that a number of randomized trials are published for one treatment, it is common practice to measure effect size with a meta-analysis. This procedure can be used when trials have similar outcome measures and case mixes, and when data are categorical or can be meaningfully interpreted categorically, such as with the Lidcombe Program example above. In 1993, the Cochrane Collaboration (http://www.cochrane.org/) was formed with the aim of producing systematic reviews in every field of medicine. In practice, the Cochrane Collaboration has gone further and produced systematic reviews in other areas of health besides medicine, including speech-language pathology.12

It is critical for clinicians to know whether an effect size has been measured reliably. Siegel (1990) applied to the field of stuttering research Sidman's (1960) contention that the surest empirical test of reliability is replication. This use of the notion of reliability pertains to external validity. In other words, reliability ultimately deals with how well an effect size for a given treatment can be generalized from the study sample to the population of those who stutter and seek clinical help. Hence, an effect size for a treatment would be reliable if it were replicated by independent groups of researchers.13 For speech-language pathologists, a replicated treatment effect increases confidence that the treatment might work with their clients. As expressed by Hegde (2007), "until other clinicians replicate the study by using different clients, external validity of the treatment procedure remains doubtful" (p. 20).

To date, the majority of clinical trials about stuttering are nonrandomized (Hegde, 2007), and thus it is rarely possible to calculate a valid effect size for a given treatment in comparison to a control condition. Until a sufficient number of randomized trials are published to support valid meta-analyses of the efficacy of treatments for stuttering, it is tempting to consider the apparent effect sizes in nonrandomized trials as a guide to the merits of treatments. However, it is well established in medical research that nonrandomized trials overestimate the efficacy of treatments (Kunz & Oxman, 1998), and this is no less true of stuttering research.

The results of nonrandomized trials of the Lidcombe Program with preschoolers serve as an example of the caution needed when evaluating nonrandomized clinical trials. Stuttering reductions with the treatment assessed in nonrandomized trials is above 90% (e.g., Miller & Guitar, 2006; Onslow, Andrews, & Lincoln, 1994; Rousseau, Packman, Onslow, Harrison, & Jones, 2007). However, the Jones et al. (2005) RCT provided a much more modest estimate when taking account of natural recovery in a control group (see below). On balance, then, considering their unreliability, the extent of stuttering reductions in nonrandomized clinical trials of stuttering treatment should be disregarded. Positive reports from such trials should be considered simply as such—no more and no less—and no attempt to establish the magnitude of the

health outcome from them should be made. Only in the case of RCTs do we recommend that clinicians take note of some index of effect size (see below).

Given that effect size estimation is not likely to be accurate in nonrandomized trials, our methods to evaluate clinical trials do not incorporate metrics to assess reliability of observations in nonrandomized trials, in other words, reliability in the common sense of internal validity. The application of such metrics is unlikely to overcome the questionable internal validity of the estimates of effect sizes in nonrandomized trials.14 This is a departure from traditional approaches recommended in evaluating stuttering treatment research. Additionally, we recommend that such metrics to assess reliability of observations not be incorporated in assessment of effect size in randomized clinical trials. The reason for this recommendation relates to the notion presented above that the reliability of observations is not a matter of internal validity but external validity. Even if a clinician applied a detailed metric to assess the reliability of observations in randomized trials, the prime indication of reliability is whether the presence of an effect and its size had been replicated.15

To return to the Lidcombe Program example, the meaning of a single finding of an odds ratio of 7.7 in an RCT of this treatment provides a conundrum for clinicians who seek to establish an evidence base to their treatments for preschoolers. Several RCTs of the Lidcombe Program, and a meta-analysis of those RCTs, are needed to establish a reasonably trustworthy—that is, an externally valid or generalizable—estimate of effect size. So clinicians may well be cautious in concluding anything definitive about the merits of the treatment.16 Even if, in the future, such meta-analysis confirms the presence of an effect and estimates its size, the issue is far from resolved, for meta-analyses of RCTs pertain only to treatment efficacy, not treatment effectiveness. Clearly, it is a critical issue whether the Lidcombe Program produces clinically important effect sizes in the "real world" (Onslow & Yaruss, 2007) of professional clinicians outside the realms of the scientific environment of the RCT. If we need to wait a while for the efficacy of the Lidcombe Program to be shown by meta-analysis, we may need to wait even longer for its effectiveness to be shown.

The confidence of knowledge about an effect size relates directly to the number of participants in trials. Putting it simply, evidence from 200 participants that there is an effect is better than evidence from 5 participants, because 200 is a better sample of the target population than 5. All clinical trials evidence is not equal.17 Consequently, when single-subject experiments meet our definition of a clinical trial, they are accorded a low level of evidence in our taxonomy.

Randomization and Effect Size in Three Phases of Clinical Trials of Stuttering Treatment

To establish our taxonomy of clinical trials of stuttering treatment development, we incorporate the principles of randomization and effect size, as discussed above, into a simple process of allocating treatments into one of three phases. Phase of evidence is a somewhat loosely applied concept for categorizing the stage of development of a treatment. The schema is most commonly applied to pharmacological treatments, but it is now commonly also applied to the development of other types of health care methods, as described, for example, by Herson (1984), Piantadosi (1997), and Pocock (1983). Recently Robey (2005) drew attention to its potential value within the discipline of speech-language pathology. In the following, those sources are drawn from in applying the schema—with some flexibilities of interpretation—to the development of stuttering treatments. It is the case that in some disciplines only the advanced level of evidence (Phase III, see below) would be considered acceptable: for example, De Angelis et al. (2007). However, as will be apparent, in the field of stuttering research, the bulk of speech-language pathology clinical trials evidence is available at an earlier phase of clinical trials development.

Phase I Trials

Phase I clinical trials are preliminary investigations of a new treatment, conducted on only a few participants who may be volunteers. The number of participants is determined arbitrarily. Phase I trials may be prompted by theoretical or empirical sources such as laboratory evidence indicating a potentially worthwhile treatment. The primary considerations in such trials are safety issues, and viability from the perspective of the client and the service provider. During Phase I trials, participants are carefully monitored, and compliance with the treatment is determined. Development of a treatment protocol may begin or be completed during Phase I. Some preliminary evidence of the existence of a treatment effect is required in order for treatment development to proceed to Phase II. In the case of pharmacological treatment development, Phase I is when the requisite dose of medication is determined in preparation for Phase II trialing. For stuttering treatments, an analog of "dose" would be the frequency and duration of treatment, such as 1 hr of treatment at weekly intervals for 20 weeks. It is important to have that information established, and ideally manualized, so that the results of Phase II trials can be properly interpreted.

For a Phase I trial of a stuttering treatment, participant numbers would be usually fewer than 10. It would normally be unethical to subject any greater number of participants to an unknown treatment in a trial. Phase I reports of innovative treatment methods might promote more interest and anticipation of Phase II trialing. Examples would include the Trajkovski, Andrews, O'Brian, Onslow, and Packman (2006) case study report of nonprogrammed rhythmic stimulation with a preschooler, and R. J. Ingham et al.'s (2001) single-subject experiments with the modification of phonation interval procedure, and James's

(1981) report of a self-imposed time-out treatment program. Phase I reports that are not followed up with Phase II treatment developments after a long period would normally cause clinicians to become less inclined to consider them meritorious.

Phase II Trials

Safety and viability, and careful monitoring of participants, remain considerations during Phase II trials. However, the prime considerations during this stage of treatment development are to establish estimates of how many participants will respond to the treatment. Phase II trials data that are unconvincing on either of these points potentially indicate that treatment development should be stopped at this phase. In a worse-case scenario, a treatment that produced a small treatment effect in few participants after a long period of treatment would be abandoned without question. Phase II trials would normally involve more than 10 participants.

As is the case with Phase I trials, the selection of sample size is often fixed with somewhat arbitrary methods for the purposes of making the above observations. This simple approach involves the observation of the proportion of "responders" within a fixed sample size (Herson, 1984; Piantadosi, 1997), as was the case with Lewis et al. (2008). Evidence that a Phase III trial is warranted is based on mathematical inference of the defined number of "responders." Alternatively, participant numbers can be set a priori in Phase II trials using specialized methods. Encouraging results of Phase II trials data give a green flag for moving treatment development into Phase III.

The number of participants in a Phase II trial of a stuttering treatment will normally be greater than 10. Replication of results by independent groups is not essential but adds weight to the generalization that is possible about the potential efficacy of the treatment. An example of a treatment at Phase II development at the time of writing is multiday intensive speech restructuring with programmed instruction (e.g., Andrews & Feyer, 1985; Block, Onslow, Packman, & Dacakis, 2006; Boberg & Kully, 1994; Howie et al., 1981; Onslow et al., 1996).

Phase III Trials

Phase III evidence is the so-called "gold standard" clinical trials evidence for a treatment. Participant numbers are not determined arbitrarily but are based on the results of Phase II trials, which enable power calculations. Participant numbers are normally substantial and range from fewer than 100 in cases where a large effect size is thought to be present to thousands of participants in cases where a small effect size is thought to be present. Phase III trials provide robust estimates of effect sizes.

Phase III trials compare a treatment to a control group or another treatment or treatments, where sufficient participant numbers allow a mathematical estimate of effect size. Replication of randomized trials adds weight to the generalization that is possible about the potential value of the treatment. In the most convincing interpretation of such replications, an odds ratio can be established from meta-analysis (see above). An example of a Phase III trial of a stuttering treatment would be the RCT reported by Jones et al. (2005). Randomized trials with insufficient participant numbers to provide mathematical estimates of effect size are relegated to Phase II status. An example would be Lewis et al. (2008).

Normally, Phase III evidence would not be assembled without underpinning Phase II evidence, because it would not be ethical to randomize participants to no-treatment control groups or potentially poor comparison groups. This would not be the case in trials where a best practice treatment is given to all participants and supplemented with an experimental add-on for one group, as occurred for example with the Menzies et al. (in press) trial of a cognitive behavior therapy add-on to speech restructuring.

 Identifying and Evaluating Clinical Trials of Stuttering Treatment: Results

Articles Eligible for Inclusion to Circa 2008

In addition to conforming to the definition of a clinical trial of a stuttering treatment presented above, the criteria for inclusion of a research article in the present review are that the article (a) was published in or after 1965, (b) was published in a generally accessible format, and (c) was available in English. Articles for review were identified through online searching of the Web of Science, CINAHL (Cumulative Index to Nursing and Allied Health Literature), Journals@Ovid, Medline, Proquest, and PsycINFO databases, and the authors' collections of textbooks and treatment manuals. It was not deemed necessary for an article to be published in a peer-reviewed journal in order to be selected for review. Book chapters, published conference proceedings, and treatment manuals were eligible for consideration providing sufficient information was available to determine whether the criteria for a clinical trial were met. Unpublished conference papers and posters, and unpublished dissertations, were not reviewed.

Trials were classified into four age groups as follows: preschool (<6 years), school-age (7–12 years), adolescent (13–17 years), and adult (>17 years). In cases where a trial reported data for participants in more than one age group, and the data for the different age groups were identifiable, the trial was reported in more than one age group. In cases where a trial reported data for the majority of participants in one age group, and those few outside the age group could not be identified, the trial was reported in the age group for the majority of its

participants. For example, participants in the Craig et al. (1996) trial contained some adolescents whose data could not be individually identified. However, in the case of the EMG arm of this trial, the mean age of participants was 11 years, with a standard deviation of 1.7 years, indicating that the majority of participants were in the school-age range. Hence, the trial was placed in the school-age group.

Tables 1–4 present the results of these searches and record the phase of trial development to which each trial can be allocated. The summary of that material is presented below. Self-imposed time-out, gradual increase in length and complexity of utterance, and the Lidcombe Program were categorized as verbal response contingent treatments.

TABLE 1 Clinical trials for adults (>17 years) that meet the definition of a clinical trial and whether each trial represents a Phase I, Phase II, or Phase III development.

TABLE 2 Clinical trials for adolescents (13–17 years old) that meet the definition of a clinical trial and whether each trial represents a Phase I, Phase II, or Phase III development.

TABLE 3 Clinical trials for school-age children (7–12 years old) that meet the definition of a clinical trial and whether each trial represents a Phase I, Phase II, or Phase III development.

TABLE 4 Clinical trials for preschool children (<6 years old) that meet the definition of a clinical trial and whether each trial represents a Phase I, Phase II, or Phase III development.

Reported Outcomes

All trials reported to date have incorporated a measure of stuttering severity as a primary outcome.18 No trial has incorporated a nonspeech measure as a primary outcome, with the exception of the Menzies et al. (in press) trial, which included psychological measures to assess outcomes of a cognitive behavior therapy package.

Trials of Treatments for Adults

There are replicated Phase II clinical trials of multiday, intensive, programmed, speech-restructuring treatments to suggest that an effect may be procured with such treatments. As noted by Bothe, Davidow, Bramlett, and Ingham (2006), the role of the speech pattern in the apparent success of these treatments is not clear. Within treatment programs, there are many clinical models for presenting speech restructuring, and many supplements to the basic speech pattern used. There is a Phase II clinical trial of speech restructuring presented in telehealth format. Replicated Phase I trials of verbal response contingent stimulation (self-

imposed time-out) have been published, as well as an unreplicated Phase II trial of this procedure. However, without the publication of a Phase III trial, the size of effects associated with any of these treatments is unknown. Of particular interest is unreplicated Phase II trials evidence in favor of combined speech restructuring and verbal response contingent treatment. An unreplicated Phase I trial of a machine-driven modification of phonation intervals treatment has been published, as has a Phase III trial of regulated breathing. These two treatments warrant interest. There has been one clinical trial of an auditory feedback device that showed no effect.19

Trials of Treatments for Adolescents

There have been replications of multiday, intensive, programmed, speech-restructuring treatments at Phase I, with an unreplicated Phase II clinical trial. No other treatments have been replicated in trials. There have been unreplicated trials of time-out, regulated breathing, and EMG biofeedback that warrant interest. In the case of the latter trial, an absence of effect was reported.

Trials of Treatments for School-Age Children

The situation with multiday, intensive, programmed, speech-restructuring treatments is similar to that described above with adolescents; there have been replications at Phase I, with an unreplicated Phase II clinical trial. There have been no other replications. There has been a Phase II failure to replicate a Phase II finding of an effect with EMG biofeedback. There have been unreplicated Phase I trials of regulated breathing and verbal response contingent stimulation (gradual increase in length and complexity of utterance), and an unreplicated Phase II trial of a verbal response contingent stimulation treatment (the Lidcombe Program). These verbal response contingent treatments warrant interest.

Trials of Treatments for Preschool Children

The bulk of clinical trials evidence is for verbal response contingent stimulation, and there are replicated Phase I and Phase II trials, and an unreplicated Phase III trial. There is Phase I and Phase II clinical trials evidence of verbal response contingent stimulation (the Lidcombe Program) presented in telehealth format, the Phase II trial being randomized and controlled. There is a Phase I report of a family-based treatment (parent–child interaction therapy; Millard, Nicholas, & Cook, 2008) that has some similarities to that reported by Yaruss et al. (2006).20 A Phase I trial of nonprogrammed syllable timed speech has been reported but, with 1 participant, warrants only passing interest at this time.

Identifying Subsequent Clinical Trials of Stuttering Treatment for Evaluation

Based on the present recommended taxonomy, the foregoing summary is an accurate and meaningful statement for clinicians about the state of clinical trials of stuttering treatment. Now it should be a simple matter for clinicians to update this information using procedures outlined below.

Step 1: Set Up an Online Publication Alert

With modern publicly accessible databases and e-mail publication alerting systems, it is a simple matter to identify a publication in the area of stuttering in a timely fashion. Such databases are continually being established and refined; however, at present we agree with Robey's (2005) recommendation of the PubMed database of the National Centre for Biotechnology Information (http://hinari-gw.who.int/whalecomwww.ncbi.nlm.nih.gov/whalecom0/entrez/query.fcgi?db=PubMed). This is a free e-mail alerting system assembled by the U.S. National Library of Medicine and the National Institutes of Health. A personalized account can be established and a search set up in a few minutes. We recommend the search string stutter* OR stammer* or dysfluen*. E-mail alerts to publications identified with such a search string can be selected to occur at regular periods; however, we recommend a weekly alert for any publications.

Step 2: Assess Whether a Publication May Be a Clinical Trial

Because the search string will identify any publication in the area of stuttering, the next step is to determine whether the publication might be a clinical trial according to the definition proposed in the present taxonomy. PubMed will e-mail the titles and abstracts of all publications, and Step 2 involves simply scanning the titles of publications. This process will quickly screen any potential clinical trials. For example, the following titles are unlikely to be trials:

Disconnection of speech-relevant brain areas in persistent developmental stuttering (Sommer, Koch, Paulus, Weiller, & Buchel, 2002)

Dichotic ear preferences of stuttering children and adults (Sommers, Brady, & Moore, 1975)

In contrast, the following titles may well constitute clinical trials:

Treating preschool children who stutter: Description and preliminary evaluation of a family-focused treatment approach (Yaruss et al., 2006)

Long-term results of an intensive treatment program for adults and adolescents who stutter (Boberg & Kully, 1994)

Experimental treatment of early stuttering: A preliminary study (Franken et al., 2005)

On many occasions, false hits will occur because of the use of the words "stutter" and "stuttering" in other disciplines. These false hits can readily be identified and discarded.

Step 3: Read the Abstract of Any Publication That Might Be a Clinical Trial

On many occasions, the PubMed abstract will confirm that the report conforms to the present definition of a clinical trial by providing information that an entire treatment was involved, at least 3 months follow-up, and that speech observations were collected from independent, beyond-clinic, conversational speech recordings. This is the case for the Boberg and Kully (1994) example above, in which case Step 4 (below) is implemented. However, in the case of the Franken et al. (2005) example, the abstract reveals that the report dealt with only a portion of two treatments, and Step 4 is not necessary. On other occasions, the abstract will leave some ambiguity about the status of the report, as occurs with the Yaruss et al. (2006) abstract, in which case Step 4 is implemented.

Step 4: Acquire and Read the Article to Confirm or Refute Its Status as a Clinical Trial of a Stuttering Treatment

The information in a report will allow a judgment of whether it warrants classification as a clinical trial of a stuttering treatment according to the present definition, and to which phase of clinical trial development it should be allocated. This involves the following 10-item checklist:

Prospective method

Presence of an outcome

An entire treatment

Pretreatment observation(s)

Follow-up observation(s) for a reported effect

Independent observations

Beyond-clinic observations

Observations based on conversation recordings

In the event that a report is consistent with that definition of a clinical trial of a stuttering treatment, it is subsequently allocated to one of the three phases of trial development, taking account of the following:

Participant numbers

Randomization

The transition from Step 3 to Step 4 is a conservative one to the extent that several reports will be read that in fact are not a clinical trial. For example, a full reading of the Yaruss et al. (2006) report shows that it is in fact a file audit and that speech observations are not based on recordings. Nonetheless, as indicated above, such file audit reports, and others, are of potential interest to clinicians and worth acquisition and reading. Additionally, in cases when readers at Step 3 identify reports that are not a clinical trial of a stuttering treatment according to the present definition, it may be considered worthwhile to proceed to Step 4 regardless. For example, even though the Franken et al. (2005) report was not a clinical trial because it involved only a portion of two treatments, the comparison would nonetheless be of potential interest.

Final Comments

The impetus for the preparation of the present taxonomy was the burden of work for clinicians in keeping up to date with information to consider during the cycle of evidence-based reasoning. Effortlessly, with publicly available search engines such as PubMed, clinicians are likely to detect the majority of published clinical trials of stuttering treatment. This is particularly the case if PubMed is supplemented by alerts from the many other search engines available. The trials found in book chapters, conference proceedings, and treatment manuals were a minor portion of the trials identified in the present article, and on balance, important trials are unlikely to be reported solely in such sources. Having detected a clinical trial of a stuttering treatment, it is a simple matter to allocate it to one of three phases of clinical trials development.

It is important to acknowledge the limits of the present taxonomy, and indeed any such taxonomy of treatment efficacy research in stuttering. Fundamentally, evidence-based practice is informed more widely than by sources of evidence alone. Whether or not a source of evidence for treatment efficacy is applied in clinical practice depends on nonscientific issues such as the needs of clients, workplace restrictions on treatment formats, and clinical expertise in treatment methods. Another limitation is, simply, that evidence-based practice is informed more widely than by evidence for treatment efficacy. Whether or not a treatment is likely to be efficacious for clients is certainly a fundamental concern in evidence-based practice. However, clinicians need to be informed about a range of research findings that have an impact on clinical practice. These include changing perspectives on the nature of the disorder, the developmental course of the untreated disorder, the risk factors for developing the disorder, and the psychological effects of the disorder. How clinicians might learn about those findings is a topic for another discussion altogether.

Perhaps the ultimate caveat to any taxonomy of the output of science is the limits of science itself. Siegel's (1987) seminal article applied this common epidemiological thought to science in communication disorders by pointing out that there is much in the conduct of our discipline's clinical practice that is beyond the realm of science, such as logic, social and personal values, and the self-evident. Particularly with regard to the latter domain, health care will go awry if science is sourced for guidance (Smith & Pell, 2003).

Footnotes

1 Throughout this article, the terms "efficacy" and "efficacious" are used in their strictest sense and not used interchangeably with "effectiveness" and "effective," as sometimes occurs.

2 Bothe, Davidow, Bramlett, and Ingham (2006) and Bothe, Davidow, Bramlett, Franic, and Ingham (2006) developed separate taxonomies for nonpharmacological and pharmacological treatments. However, because our purpose is oriented to practitioners of speech-language pathology, we are concerned only with the former.

3 The clinician consumer of clinical trials research needs to be vigilant for the misleading potential of the post hoc statistical analysis: the possibility of being convinced that statistical evidence has been discovered for the efficacy of a treatment without a primary hypothesis and associated primary outcome. The distinction between the capacity of exploratory clinical research to inform rather than mislead clinicians can be subtle. This is highlighted in the Blomgren, Roy, Callister, and Merrill (2005) report dealing with the Successful Stuttering Modification Program (SSMP), which contained 14 outcome measures for 19 participants. Significant changes in 4 of these outcomes from pretreatment to posttreatment prompted a conclusion that "findings indicate that the SSMP, at its core, is an anxiolytic treatment" (p. 520). Should a clinician reading this report form the view that preliminary evidence for an anxiolytic treatment for stuttering has been discovered or that the value of such a treatment has been demonstrated in a clinical trial? Those two views will have different impacts on clinical practices.

4 We use the term "follow-up" rather than "posttreatment" to highlight the passage of time.

5 Quite possibly, many noneffects have been found but not written up for publication, in what has been referred to as the "file drawer effect" (Rosenthal, 1979). In the event that such reports are submitted for publication, they are less likely to be accepted for publication in a peer-reviewed journal than positive findings. Indeed, the present review found few published clinical trials that reported a null result. In an extreme scenario, it would be extremely misleading if, for every one nonrandomized published report of a popular treatment showing an effect, a trial had been conducted and found no effect but yet was not published. A recent development is for journals to require registration of the trial in a clinical trials register. One of the reasons for such registers is to provide a public record of all trials, regardless of their results. This potentially would counteract the file drawer effect.

6 We chose 3 months because short follow-up periods might be of some value in preliminary reports that demonstrate the effects of a new treatment. For a traditional treatment that requires many clinical hours, a longish follow-up of 12 months might be needed to determine whether a worthwhile benefit resulted from such clinical expenditure. However, one appeal of auditory devices for alleviating stuttering is that they obviate the traditional notion of treatment time; any treatment effects would begin at the time the speaker begins to wear such a device. It would be worth knowing whether clinical benefits were attained as soon as the speaker began to wear such a device.

7 We assume that an independent observer in a trial would be blinded to the treatment status of participants.

8 Randomization is not a complete solution to the problem of bias, as is often implied with the use of the term "gold standard" in reference to this methodology (Hegde, 2007). Hegde points out that randomization in clinical trials is random allocation, not random selection of participants. The biases in selecting participants for random allocation include the fact that (a) inevitably, to some extent, participants self-select for inclusion in trials; (b) those included in trials are required to sign informed consent; and (c) those who conduct the trial identify the pool of potential participants.

9 Another source of bias when making repeated observations of participants is associated with sampling error rather than actual change in severity of stuttering. Extreme scores, either high or low, are likely to be followed by less extreme scores closer to the actual mean stuttering severity of the participant. Hence, when measuring the severity of a group on two occasions, the ordering of individuals from lowest to highest severity will be different on the two occasions because of measurement error. Therefore, an apparently severe subgroup on the first measurement occasion will not be the same as an

apparently severe subgroup on the second measurement occasion. Therefore, the first group who is apparently most severely affected will have lower stuttering severity measured on the second occasion (see Zhang & Tomblin, 2003.) This is called regression to the mean, and one way it can be accounted for is with a randomized trial design (Barnett, van der Pols, & Dobson, 2005).

10 In reality, there is no need for this to occur in nonrandomized trials, or in any trials for that matter. There is no reason why Onslow et al., or others who have reported nonrandomized trials, could not have used the "intention to treat" principle to provide a better estimate of the efficacy of their treatments. Dropouts are handled so that trial outcomes are reported including those participants. Less commonly, in what is known as a cross-over, when a participant moves to a different arm of the trial, outcomes for that participant are presented for the arm originally allocated. Another intention to treat approach is "last observation carried forward," which means that when a participant drops out, the last available data point for that participant is carried forward to the final assessment or assessments. In other words, if a participant is assessed pretreatment at 25% syllables stuttered (%SS) and then drops out before receiving treatment, 25%SS is deemed to be that participant's follow-up stuttering severity at the posttreatment assessment or assessments.

11 There are various other methods of assessing effect size in randomized clinical trials. All methods used involve the calculation of a weighted average of the summary statistics from the included trials. The weights are normally based on the amount of information each study contains, which is sometimes estimated by the sample size but more commonly by the inverse of the variance in the data. The standard error of the weighted average is also estimated so that a confidence interval and p value can be calculated. More details on statistical methods for meta-analysis can be found in Whitehead (2002).

12 An example of a systematic review and meta-analysis that has been published by the Cochrane Collaboration is of oral sumatriptan for treatment of acute migraine (McCrory & Gray, 2004). The objective of the review was "to describe and assess the evidence from randomised controlled trials (RCTs) concerning the efficacy and tolerability of oral sumatriptan for the treatment of a single acute attack of migraine in adults" (McCrory & Gray, 2004, p. 1). One of the primary efficacy outcomes assessed in the review was pain-free response: complete resolution of headache pain. This meta-analysis results in an estimated treatment effect for pain-free response of 4.19 odds ratio, with a 95% confidence interval of 3.13 to 5.61, in favor of 100 mg of sumatriptan compared with placebo. In other words, the odds of complete resolution of headache pain at 2 hr is more than four times greater with 100 mg of sumatriptan compared with placebo. The systematic review concluded that "oral sumatriptan has been shown to be an effective drug for the treatment of a single acute attack of

migraine. It is well tolerated, though minor adverse events were not uncommon in the included trials" (McCrory & Gray, 2004, p 2).

13 It is important to not conflate two common meanings of the term "generalization." This term most commonly is used to mean establishing a generality about a population by means of replicated findings by independent groups of researchers. However, in clinical and single-subject experimental applications, it can also mean establishing that a response of an individual is present in a range of environments. Many single-subject experimental designs take account of the latter issue, most notably multiple baseline across situations designs. For a discussion of the issue of determining generalization in single-subject experimental designs, see Kendall (1981). A clouded distinction of these two meanings of generalization can lead to fallacious arguments about the external validity of single-subject experimentation (e.g., McReynolds & Thompson, 1986).

14 Bothe, Davidow, Bramlett, and Ingham's (2006) proposed criterion interjudge data reliability of 80% or 0.80 is presumably fueled by the well-known fact that much error may occur when different observers attempt to count stuttering and stuttering-related speech events. There are issues that arise from such numerical determination of levels of interjudge reliability. In the first instance, high levels of agreement on absolute scores between judges means only that they agree. However, they both may agree but be affected by the same source of bias. Second, correlation, of itself, is not an adequate reliability index, because two observers can have completely different scores and have a high correlation. According to a reanalysis of the Kully and Boberg (1988) data by Onslow (1996), this is a possible—even likely—scenario. Hence, the only comprehensive assessment of reliability comprises an index of correlation combined with pairwise differences for both interobserver and intraobserver agreement (Onslow, Adams, & Ingham, 1992), but there is no sign to date of the acceptance by stuttering treatment researchers of this fact.

15 We do not mean that reliability for the sake of internal validity is unimportant scientifically. To the contrary, in doing the science of clinical trials, it is critical to maximize internal validity by doing all to ensure that observations are reliable. But our point is that for the present purposes of establishing methods for clinicians to evaluate clinical trials, reliability from the perspective of internal validity is not a prime concern.

16 In the interim, Onslow, Jones, Menzies, O'Brian, and Packman (2008) provided a preliminary estimate of effect size for the Lidcombe Program with a meta-analysis. They added to the Jones et al. (2005) and Lewis, Onslow, Packman, Jones, and Simpson (2008) randomized trials two randomized controlled experiments involving the Lidcombe Program (Harris et al., 2002; Lattermann et al., 2008). These experiments did not deal with a complete

treatment, but compared, respectively, 12-week and 16-week portions of the Lidcombe Program to no treatment. Onslow et al. (2008) combined the data from these two studies and concluded that children treated with the Lidcombe Program for varying periods had higher odds of attaining minimal stuttering (<1.0%SS). The odds ratio was 7.5 (95% confidence interval = 2.7–20.9, p < .001).

17 On this point we join others (Bloodstein, 1995; Siegel & Young, 1987), who are apparently the minority, in departing from some other views of the matter (Bothe, Davidow, Bramlett, & Ingham, 2006; Davidow et al., 2006; Hegde, 2007; McReynolds & Thompson, 1986) that accord particular emphasis on the value of single-subject design experiments. We agree with contentions about the internal validity of such methodology. And we agree with Bothe, Davidow, Bramlett, and Ingham (2006) that single-subject designs are part of the tradition of stuttering treatment research. But there is no need for speech-language pathology research traditions to override the basic principles of clinical trials that direct the development of human health care systems. Single-subject experiments do allow a certain kind of strong, scientific reasoning, based on their tight internal control. We agree with arguments that the group design is not the ultimate alternative to establishing external validity (Costello, 1979; Hegde, 2007; Siegel & Ingham, 1987), and that replicated single-subject experiments do contribute to establishing external validity (Hegde, 2007; McReynolds & Thompson, 1986). However single-subject experiments do not provide adequate participant numbers or the mathematical apparatus to allow for generalized statements about effect sizes that can be applied to treated populations.

18 In the majority of cases in the treatment efficacy literature on stuttering, authors make it clear that there are one or two primary outcomes of major interest. For example, Langevin et al. (2006) stated that their report was to evaluate the "effectiveness of the ISTAR Comprehensive Stuttering Program (CSP) within and across client groups from the Netherlands and Canada" (p. 229). Langevin et al.'s description of the treatment process goals was "acquisition of fluency and cognitive behavioral skills" (p. 235). Under their "treatment outcome measures" listed on page 236, they state that "the dependent speech measures were percent syllables stuttered (% SS) and syllables spoken per minute (SPM)," and on page 237 they state that speech naturalness ratings were collected "to give context for the interpretation of" those measures. Last of all, a collection of five "self report measures" are presented on page 236. Considering that "acquisition of fluency" is the first stated goal of the program, and that Langevin et al. present the %SS and SPM measures first in their report, it appears that these are intended as primary outcomes.

19 At first impression, the report by O'Donnell, Armson, and Kiefte (2008) appears to be a clinical trial of an altered auditory feedback device according to the definition presented here. However, during a weekly experimenter phone call to some participants at 16 weeks posttreatment, while they were wearing the device, "participants were asked about their subjective impressions of benefit and gave a description of their experiences wearing the device" (p. 104). For this reason, we regarded this assessment as not independent of treatment, and hence the report to not be a clinical trial according to the present definition.

20 Millard et al. reported 6 cases, but the nature of their data precluded assessment of whether an effect was present. They provided us with pretreatment and follow-up %SS data. We established that the 6 children had a mean stuttering reduction of 64% with considerable variability (S1 = 66%; S2 = 10%; S3 = 74%; S4 = 97%; S5 = 95%; S6 = 44%).

References

Ambrose, N. G., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. Journal of Speech, Language, and Hearing Research, 42, 895–909.[Abstract/Free Full Text]

Andrews, G., & Feyer, A. M. (1985). Does behavior therapy still work when the experimenters depart? An analysis of a behavioral treatment program for stuttering. Behavior Modification, 9, 443–457.[Abstract]

Andrews, G., Guitar, B., & Howie, P. (1980). Meta-analysis of the effects of stuttering treatment. Journal of Speech and Hearing Disorders, 45, 287–307.[Abstract/Free Full Text]

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. International Journal of Epidemiology, 34, 215–220.[Abstract/Free Full Text]

Block, S., Onslow, M., Packman, A., & Dacakis, G. (2006). Connecting stuttering management and measurement: IV. Predictors of outcome for a behavioural treatment for stuttering. International Journal of Language and Communication Disorders, 41, 395–406.[CrossRef]

Block, S., Onslow, M., Packman, A., Gray, B., & Dacakis, G. (2005). Treatment of chronic stuttering: Outcomes from a student training clinic. International Journal of Language and Communication Disorders, 40, 455–466.[CrossRef]

Block, S., Onslow, M., Roberts, R., & White, S. (2004). Control of stuttering with EMG feedback. Advances in Speech Language Pathology, 6(2), 100–106.[CrossRef]

Blomgren, M., Roy, N., Callister, T., & Merrill, R. M. (2005). Intensive stuttering modification therapy: A multidimensional assessment of treatment outcomes. Journal of Speech, Language, and Hearing Research, 48, 509–523.[Abstract/Free Full Text]

Bloodstein, O. (1995). A handbook on stuttering (5th ed.). San Diego, CA: Singular.

Boberg, E. (1981). Maintenance of fluency: An experimental program. In E. Boberg (Ed.), Maintenance of fluency (pp. 71–111). New York: Elsevier.

Boberg, E., & Kully, D. (1994). Long-term results of an intensive treatment program for adults and adolescents who stutter. Journal of Speech and Hearing Research, 37, 1050–1059.[ISI][Medline]

Bothe, A. K., Davidow, J. H., Bramlett, R. E., Franic, D. M., & Ingham, R. J. (2006). Stuttering treatment research 1970–2005: II. Systematic review incorporating trial quality assessment of pharmacological approaches. American Journal of Speech-Language Pathology, 15, 342–352.

Bothe, A. K., Davidow, J. H., Bramlett, R. E., & Ingham, R. J. (2006). Stuttering treatment research 1970–2005: I. Systematic review incorporating trial quality assessment of behavioral, cognitive, and related approaches. American Journal of Speech-Language Pathology, 15, 321–341.

Brundage, S. B., Bothe, A. K., Lengeling, A. N., & Evans, J. J. (2006). Comparing judgments of stuttering made by students, clinicians, and highly experienced judges. Journal of Fluency Disorders, 31, 271–283.[CrossRef][ISI][Medline]

The CONSORT Group. (2008). CONSORT statement. Retrieved January 15, 2008, from www.consort-statement.org/?o=1001

Conture, E. G., & Guitar, B. E. (1993). Evaluating efficacy of treatment of stuttering— school-age children. Journal of Fluency Disorders, 18, 253–287.[CrossRef][ISI]

Costello, J. (1979). Clinicians and researchers: A necessary dichotomy? Journal of the National Student Speech and Hearing Association, 7, 6–26.

Craig, A. R., & Hancock, K. (1995). Self-reported factors related to relapse following treatment for stuttering. Australian Journal of Human Communication Disorders, 23, 48–60.

Craig, A., Hancock, K., Chang, E., McCready, C., Shepley, A., McCaul, A., et al. (1996). A controlled clinical trial for stuttering in persons aged 9 to 14 years. Journal of Speech and Hearing Research, 39, 808–826.[ISI][Medline]

Curlee, R. F. (1993). Evaluating treatment efficacy for adults—assessment of stuttering disability. Journal of Fluency Disorders, 18, 319–331.[CrossRef][ISI]

Curlee, R., & Yairi, E. (1998). Treatment of early childhood stuttering: Advances and research needs. American Journal of Speech-Language Pathology, 7(3), 20–26.

Davidow, J. H., Bothe, A. K., & Bramlett, R. E. (2006). The stuttering treatment research evaluation and assessment tool (STREAT): Evaluating treatment research as part of evidence-based practice. American Journal of Speech-Language Pathology, 15, 126–141.

De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C. H., Hoey, J., Horton, R., et al. (2007). Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. Retrieved January 15, 2008, from http://www.icmje.org/clin_trialup.htm

De Klinder, M., & Boelens, H. (1998). Habit-reversal treatment for children's stuttering—assessment in three settings. Journal of Behavior Therapy and Experimental Psychiatry, 29, 261–265.[CrossRef][ISI][Medline]

Franken, M. C. J., Kielstra-Van der Schalka, C. J., & Boelens, H. (2005). Experimental treatment of early stuttering: A preliminary study. Journal of Fluency Disorders, 30, 189–199.[CrossRef][ISI][Medline]

Hancock, K., Craig, A., McCready, C., McCaul, A., Costello, D., Gilmore, G., et al. (1998). Two- to six-year controlled-trial stuttering outcomes for children and adolescents. Journal of Speech, Language, and Hearing Research, 41, 1242–1252.[Abstract/Free Full Text]

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.[CrossRef][ISI][Medline]

Harrison, E., Onslow, M., Andrews, C., Packman, A., & Webber, M. (1998). Control of stuttering with prolonged speech: Development of a one-day instatement program. In A. K. Cordes & R. J. E. Ingham (Eds.), Treatment efficacy in stuttering (pp. 191–212). San Diego, CA: Singular.

Harrison, E., Wilson, L., & Onslow, M. (1999). Distance intervention for early stuttering with the Lidcombe Programme. Advances in Speech Language Pathology, 1, 31–36.

Hearne, A., Packman, A., Onslow, M., & O'Brian, S. (2008). Developing treatments for adolescents who stutter: A Phase I trial of the Camperdown Program. Language, Speech, and Hearing Services in Schools, 39, 487–497.[Abstract/Free Full Text]

Hegde, M. N. (2007). A methodological review of randomized clinical trials. Communicative Disorders Review, 1, 15–36.

Herson, J. (1984). Monitoring data logistics with a PC. Controlled Clinical Trials, 5, 308.

Hewat, S., O'Brian, S., Onslow, M., & Packman, A. (2001). Control of chronic stuttering with self-imposed time-out: Preliminary outcome data. Papers from the 2nd Asia Pacific Conference on Speech, Language and Hearing, Part IV. Asia Pacific Journal of Speech, Language & Hearing, 6(2), 97–102.

Hewat, S., Onslow, M., Packman, A., & O'Brian, S. (2006). A Phase II clinical trial of self-imposed time-out treatment for stuttering in adults and adolescents. Disability and Rehabilitation, 28(1), 33–42.[CrossRef][ISI][Medline]

Howie, P. M., Tanner, S., & Andrews, G. (1981). Short-term and long-term outcome in an intensive treatment program for adult stutterers. Journal of Speech and Hearing Disorders, 46, 104–109.[Abstract/Free Full Text]

Huber, A., O'Brian, S., Onslow, M., & Packman, A. (2003). Results of a pilot study of EMG biofeedback for the control of stuttering in adolescents. In C. Williams & S. Leitao (Eds.), Nature, nurture, knowledge: Proceedings of the 2003 Speech Pathology Australia national conference (pp. 177–182). Melbourne, Australia: Speech Pathology Australia.

Ingham, J. C., & Riley, G. (1998). Guidelines for documentation of treatment efficacy for young children who stutter. Journal of Speech, Language, and Hearing Research, 41, 753–770.[Abstract/Free Full Text]

Ingham, R. J. (1980). Modification of maintenance and generalization during stuttering treatment. Journal of Speech and Hearing Research, 23, 732–745.[ISI][Medline]

Ingham, R. J. (1984). Stuttering and behavior therapy: Current status and experimental foundations. San Diego, CA: College Hill Press.

Ingham, R., & Andrews, G. (1973). An analysis of a token economy in stuttering therapy. Journal of Applied Behavior Analysis, 6, 219–229.[CrossRef][ISI][Medline]

Ingham, R. J., Kilgo, M., Ingham, J. C., Moglia, R., Belknap, H., & Sanchez, T. (2001). Evaluation of a stuttering treatment based on reduction of short phonation intervals. Journal of Speech, Language, and Hearing Research, 44, 1229–1244.[Abstract/Free Full Text]

James, J. E. (1981). Behavioral self-control of stuttering using time-out from speaking. Journal of Applied Behavior Analysis, 14, 25–37.[CrossRef][ISI][Medline]

James, J. E., Ricciardelli, L. A., Hunter, C. E., & Rogers, P. (1989). Relative efficacy of intensive and spaced behavioral treatment of stuttering. Behavior Modification, 13, 376–395.[Abstract]

Jones, M., Gebski, V., Onslow, M., & Packman, A. (2002). Statistical power in stuttering research: A tutorial. Journal of Speech, Language, and Hearing Research, 45, 243–255.[Abstract/Free Full Text]

Jones, M., Hearne, A., Onslow, M., Ormond, T., Williams, S., Schwarz, I., & O'Brian, S. (2008). Extended follow up of a randomised controlled trial of the Lidcombe Program of Early Stuttering Intervention. International Journal of Language and Communication Disorders, 7, 1–13.[CrossRef]

Jones, M., Onslow, M., Harrison, E., & Packman, A. (2000). Treating stuttering in young children: Predicting treatment time in the Lidcombe Program. Journal of Speech, Language, and Hearing Research, 43, 1440–1450.[Abstract/Free Full Text]

Jones, M., Onslow, M., Packman, A., Williams, S., Ormond, T., Schwarz, I., et al. (2005). Randomised controlled trial of the Lidcombe programme of early stuttering intervention. British Medical Journal, 331(7518), 659–661.[Abstract/Free Full Text]

Kendall, P. C. (1981). Assessing generalization and the single-subject strategies. Behavior Modification, 5, 307–319.[Abstract]

Kingston, M., Huber, A., Onslow, M., Jones, M., & Packman, A. (2003). Predicting treatment time with the Lidcombe Program: Replication and meta-analysis. International Journal of Language and Communication Disorders, 38, 165–177.[CrossRef]

Koushik, S., Shenker, R., Onslow, M., & Adaman, B. (2008). Follow-up of school-age children after Lidcombe Program treatment. Manuscript in preparation.

Kully, D., & Boberg, E. (1988). An investigation of interclinic agreement in the identification of fluent and stuttered syllables. Journal of Fluency Disorders, 13, 309–318.[CrossRef][ISI]

Kully, D., & Boberg, E. (1991). Therapy for school-age children. Seminars in Speech and Language, 12, 291–300.

Kunz, R., & Oxman, A. D. (1998). The unpredictability paradox: Review of empirical comparisons of randomised and non-randomised clinical trials. British Medical Journal, 317(7167), 1185–1190.[Abstract/Free Full Text]

Langevin, M., & Boberg, E. (1993). Results of an intensive stuttering therapy program. Journal of Speech-Language Pathology and Audiology, 17, 158–166.

Langevin, M., Huinck, W. J., Kully, D., Peters, H. F., Lomheim, H., & Tellers, M. (2006). A cross-cultural, long-term outcome evaluation of the ISTAR Comprehensive Stuttering Program across Dutch and Canadian adults who stutter. Journal of Fluency Disorders, 31, 229–256.[CrossRef][ISI][Medline]

Lattermann, C., Euler, H. A., & Neumann, K. A. (2008). Randomized control trial to investigate the impact of the Lidcombe Program on early stuttering in German-speaking preschoolers. Journal of Fluency Disorders, 33, 52–65.[CrossRef][ISI][Medline]

Lewis, C., Onslow, M., Packman, A., Jones, M., & Simpson, J. A. (2008). Phase II trial of telehealth delivery of the Lidcombe Program of early stuttering intervention. American Journal of Speech-Language Pathology, 17, 139–149.

Lincoln, M., & Onslow, M. (1997). Long-term outcome of an early intervention for stuttering. American Journal of Speech-Language Pathology, 6, 51–58.

Lincoln, M., Onslow, M., Lewis, C., & Wilson, L. (1996). A clinical trial of an operant treatment for school-age children who stutter. American Journal of Speech-Language Pathology, 5(2), 73–85.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. American Psychologist, 48, 1181–1209.[CrossRef][Medline]

Martin, R. (1981). Introduction and perspective: Review of published research. In E. Boberg (Ed.), Maintenance of fluency (pp. 1–30). New York: Elsevier.

Martin, R. R., Kuhl, P., & Haroldson, S. (1972). An experimental treatment with two preschool stuttering children. Journal of Speech and Hearing Research, 15, 743–752.[ISI][Medline]

McCrory, D. C., & Gray, R. N. (2004). Oral sumatriptan for acute migraine (Cochrane Review). Chichester, United Kingdom: Wiley.

McReynolds, L., & Thompson, C. (1986). Flexibility of single-subject experimental designs. Part I: Review of the basics of single-subject designs. Journal of Speech and Hearing Disorders, 51, 194–203.[Abstract/Free Full Text]

Menzies, R., O'Brian, S., Onslow, M., Packman, A., St Clare, T., & Block, S. (in press). An experimental clinical trial of a cognitive behavior therapy package for chronic stuttering. Journal of Speech, Language, and Hearing Research.

Millard, S. K., Nicholas, A., & Cook, F. M. (2008). Is parent-child interaction therapy effective in reducing stuttering? Journal of Speech, Language, and Hearing Research, 51, 636–650.[Abstract/Free Full Text]

Miller, B., & Guitar, B. (2006, November). Long term outcome of the Lidcombe Programof early stuttering intervention. Poster presented at the annual convention of the American Speech-Language-Hearing Association, Miami.

O'Brian, S., Onslow, M., Cream, A., & Packman, A. (2003). The Camperdown Program: Outcomes of a new prolonged-speech treatment model. Journal of Speech, Language, and Hearing Research, 46, 933–946.[Abstract/Free Full Text]

O'Brian, S., Packman, A., & Onslow, M. (2008). Telehealth delivery of the Camperdown Program for adults who stutter. Journal of Speech, Language, and Hearing Research, 51, 1–12.

O'Donnell, J. J., Armson, J., & Kiefte, M. (2008). The effectiveness of SpeechEasy during situations of daily living. Journal of Fluency Disorders, 33, 99–119.[CrossRef][ISI][Medline]

Onslow, M. (1995). A picture is worth more than any words [Comment]. Journal of Speech and Hearing Research, 38, 586–588.

Onslow, M. (1996). Behavioral management of stuttering (2nd ed.). San Diego, CA: Singular.

Onslow, M., Adams, R., & Ingham, R. (1992). Reliability of speech naturalness ratings of stuttered speech during treatment. Journal of Speech and Hearing Research, 35, 994–1001.[ISI][Medline]

Onslow, M., Andrews, C., & Lincoln, M. (1994). A control-experimental trial of an operant treatment for early stuttering. Journal of Speech and Hearing Research, 37, 1244–1259.[ISI][Medline]

Onslow, M., Costa, L., Andrews, C., Harrison, E., & Packman, A. (1996). Speech outcomes of a prolonged-speech treatment for stuttering. Journal of Speech and Hearing Research, 39, 734–749.[ISI][Medline]

Onslow, M., Costa, L., & Rue, S. (1990). Direct early intervention with stuttering: Some preliminary data. Journal of Speech and Hearing Disorders, 55, 405–416.[Abstract/Free Full Text]

Onslow, M., Jones, M., Menzies, R., O'Brian, S., & Packman, A. (2008). Meta analysis of Lidcombe Program treatment effects. Manuscript in preparation.

Onslow, M., O'Brian, S., Packman, A., & Rousseau, I. (2004). Long term follow up of speech outcomes for a prolonged-speech treatment for stuttering: The effects of paradox on stuttering treatment research. In A. K. Bothe (Ed.), Evidence-based treatment of stuttering: Empirical issues and clinical implications (pp. 231–244). Mahwah, NJ: Erlbaum.

Onslow, M., & Packman, A. (1999). Treatment recovery and spontaneous recovery from early stuttering: The need for consistent methods in collecting and interpreting data. Journal of Speech, Language, and Hearing Research, 42, 398–402.[Free Full Text]

Onslow, M., & Yaruss, J. S. (2007). Differing perspectives on what to do with a stuttering preschooler and why. American Journal of Speech-Language Pathology, 16, 65–68.

Packman, A., Onslow, M., O'Brian, S., & Huber, A. (2004). Measuring the outcomes of behavioural stuttering treatments: What are our data telling us? In A. K. Bothe (Ed.), Evidence-based treatment of stuttering: Empirical issues and clinical implications (pp. 245–258). Mahwah, NJ: Erlbaum.

Perkins, W. (1981). Measurement and maintenance of fluency. New York: Elsevier.

Perkins, W. H., Rudas, J., Johnson, L., Michael, W. B., & Curlee, R. F. (1974). Replacement of stuttering with normal speech. 3. Clinical effectiveness. Journal of Speech and Hearing Disorders, 39, 416–428.[Abstract/Free Full Text]

Piantadosi, S. (1997). Clinical trials: A methodological perspective. New York: Wiley.

Pocock, S. J. (1983). Clinical trials. Chichester, United Kingdom: Wiley.

Pollard, R., Ellis, J. B., Finan, D., & Ramig, P. R. (in press). Effects of the SpeechEasy on objective and perceived aspects of stuttering: a six-month, Phase I clinical trial in naturalistic environments. Journal of Speech, Language, and Hearing Research.

Reed, C., & Godden, A. (1977). An experimental treatment using verbal punishment with two preschool stutterers. Journal of Fluency Disorders, 2, 225–233.[CrossRef][ISI]

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.[Abstract/Free Full Text]

Riley, J., & Riley, G. (1999). Speech motor training. In M. Onslow & A. Packman (Eds.), The handbook of early stuttering intervention. San Diego, CA: Singular.

Robey, R. (2005, May 24). An introduction to clinical trials. The ASHA Leader, 6–23.

Rosenthal, R. (1979). The "file drawer" problem and tolerance for null results. Psychological Bulletin, 86, 638–641.[CrossRef][ISI]

Rousseau, I., Packman, A., Onslow, M., Harrison, L., & Jones, M. (2007). Language, phonology, and treatment time in the Lidcombe Program: A prospective study in a Phase II trial. Journal of Communication Disorders, 40, 382–397.[CrossRef][ISI][Medline]

Ryan, B. P., & Van Kirk Ryan, B. (1983). Programmed stuttering therapy for children: Comparison of four establishment programs. Journal of Fluency Disorders, 8, 291–322.[CrossRef][ISI]

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

Saint Laurent, L., & Ladouceur, R. (1987). Massed versus distributed application of the regulated-breathing method for stutterers and its long-term effect. Behavior Therapy, 18, 38–50.[CrossRef][ISI]

Sidman, M. (1960). Tactics of scientific research. New York: Basic Books.

Siegel, G. (1987). The limits of science in communication disorders. Journal of Speech and Hearing Disorders, 52, 306–312.[Abstract/Free Full Text]

Siegel, G. (1990). Conference on treatment efficacy: Concluding remarks. In L. Olswang, C. Thompson, S. Warren, & N. Minghetti (Eds.), Treatment efficacy research in communication disorders (pp. 233–237). Rockville, MD: American Speech-Language-Hearing Foundation.

Siegel, G. M., & Ingham, R. J. (1987). Theory and science in communication disorders. Journal of Speech and Hearing Disorders, 52, 99–104.[Abstract/Free Full Text]

Siegel, G., & Young, M. (1987). Group designs in clinical research. Journal of Speech and Hearing Disorders, 52, 194–199.[Abstract/Free Full Text]

Smith, G. C. S., & Pell, J. P. (2003). Parachute use to prevent death and major trauma related to gravitational challenge: Systematic review of randomised controlled trials. British Medical Journal, 327, 1451–1459.[Free Full Text]

Sommer, M., Koch, M. A., Paulus, W., Weiller, C., & Buchel, C. (2002). Disconnection of speech-relevant brain areas in persistent developmental stuttering. Lancet, 360(9330), 380–383.[CrossRef][ISI][Medline]

Sommers, R. K., Brady, W. A., & Moore, W. H., Jr. (1975). Dichotic ear preferences of stuttering children and adults. Perceptual and Motor Skills, 41, 931–938.[ISI][Medline]

Starkweather, C. W. (1993). Issues in the efficacy of treatment for fluency disorders. Journal of Fluency Disorders, 18, 151–168.[CrossRef][ISI]

Starkweather, C. W., & Gottwald, S. R. (1993). A pilot study of relations among measures obtained at intake and discharge in a program of prevention and early intervention for stuttering. American Journal of Speech-Language Pathology, 2, 51–58.

St. Louis, K., & Westbrook, J. (1987). The effectiveness of treatment for stuttering. In L. Rustin, H. Purser, & D. Rowley (Eds.), Progress in the treatment of fluency disorders (pp. 235–257). London: Taylor & Francis.

Stokes, T., & Baer, D. (1977). An implicit technology of generalization. Journal of Applied Behavior Analysis, 10, 349–367.[CrossRef][ISI][Medline]

Teesson, K., Packman, A., & Onslow, M. (2003). The Lidcombe behavioral data language of stuttering. Journal of Speech, Language, and Hearing Research, 46, 1009–1015.[Abstract/Free Full Text]

Thomas, C., & Howell, P. (2001). Assessing efficacy of stuttering treatments. Journal of Fluency Disorders, 26, 311–333.[CrossRef][ISI][Medline]

Trajkovski, N., Andrews, C., O'Brian, S., Onslow, M., & Packman, A. (2006). Treating stuttering in a preschool child with syllable timed speech: A case study. Behavior Change, 23(4), 270–277.[CrossRef]

Webster, R. (1980). Evolution of a target-based behavioral therapy for stuttering. Journal of Fluency Disorders, 5, 303–320.[CrossRef][ISI]

Whitehead, A. (2002). Meta-analysis of controlled clinical trials. Chichester; United Kingdom: Wiley.

Wilson, L., Onslow, M., & Lincoln, M. (2004). Telehealth adaptation of the Lidcombe Program of early stuttering intervention: Preliminary data. American Journal of Speech-Language Pathology, 13, 81–93.

Yaruss, I. S., Coleman, C., & Hammer, D. (2006). Treating preschool children who stutter: Description and preliminary evaluation of a family-focused treatment approach. Language, Speech, and Hearing Services in Schools, 37, 118–132.[Abstract/Free Full Text]

Zhang, X., & Tomblin, J. B. (2003). Explaining and controlling regression to the mean in longitudinal research designs. Journal of Speech, Language, and Hearing Research, 46, 1340–1351.[Abstract/Free Full Text]

TABLE 1 Clinical trials for adults (>17 years) that meet the definition of a clinical trial and whether each trial represents a Phase I, Phase II, or Phase III development.

Treatment	Phase	Clinical trial
Speech restructuring	Phase I	Harrison et al. (1998)
		Block et al. (2005)
	Phase II	Boberg (1981)
		Boberg & Kully (1994)
		Howie et al. (1981)
		R. Ingham & Andrews (1973)
		James et al. (1989)
		Langevin & Boberg (1993)
		Langevin et al. (2006)
		O'Brian et al. (2003)

Treatment	Phase	Reference
		O'Brian et al. (2008)
		Onslow et al. (1996)
		Onslow et al. (2004)[a]
		Perkins et al. (1974)
	Phase III	
Regulated breathing	Phase I	
	Phase II	
	Phase III	Saint-Laurent & Ladouceur (1987)
VRCS	Phase I	Hewat et al. (2001)[b]
		James (1981)[b]
	Phase II	Hewat et al. (2006)[b]
	Phase III	
Machine	Phase I	R. J. Ingham et al. (2001)
	Phase II	
	Phase III	
Cognitive behavior therapy	Phase I	
	Phase II	
	Phase III	Menzies et al. (in press)
Combined auditory feedback device and speech restructuring	Phase I	Pollard et al. (in press)[*]
	Phase II	
	Phase III	

TABLE 2 Clinical trials for adolescents (13–17 years old) that meet the definition of a clinical trial and whether each trial represents a Phase I, Phase II, or Phase III development.

Treatment	Phase	Clinical trial
Speech restructuring	Phase I	Harrison et al. (1998)
		Hearne et al. (2008) *[a]
		Langevin & Boberg (1993)
	Phase II	Boberg & Kully (1994)
	Phase III	
VRCS	Phase I	Hewat et al. (2006)
	Phase II	
	Phase III	
Regulated breathing	Phase I	De Klinder & Boelens (1998)
	Phase II	
	Phase III	
EMG biofeedback	Phase I	Huber et al. (2003)*
	Phase II	
	Phase III	

Note. EMG = electromyographic. All trials reported an effect except those marked with an asterisk.

[a]No effect was found for 1 of 3 adolescents in this trial.

TABLE 3 Clinical trials for school-age children (7–12 years old) that

meet the definition of a clinical trial and whether each trial represents a Phase I, Phase II, or Phase III development.

Treatment	Phase	Clinical trial
Speech restructuring	Phase I	Boberg & Kully (1994)
		Kully & Boberg (1991)
		Ryan & Van Kirk Ryan (1995)
	Phase II	Craig et al. (1996), Hancock et al. (1998)[a]
	Phase III	
EMG biofeedback	Phase I	
	Phase II	Block et al. (2004)*
		Craig et al. (1996), Hancock et al. (1998)[a]
	Phase III	
Regulated breathing	Phase I	De Klinder & Boelens (1998)
	Phase II	
	Phase III	
VRCS	Phase I	Ryan & Van Kirk Ryan (1983, 1995)[b]
	Phase II	Lincoln et al. (1996)[c]
	Phase III	

Note. All trials reported an effect except those marked with an asterisk.

[a]Long-term follow-up of the participants in the Craig et al. (1996) trial.

[b]Report contained data for time-out and gradual increase in length and complexity of utterance treatments.

TABLE 4 Clinical trials for preschool children (<6 years old) that meet the definition of a clinical trial and whether each trial represents a Phase I, Phase II, or Phase III development.

Treatment	Phase	Clinical trial
VRCS	Phase I	Harrison et al. (1999)
		Martin et al. (1972)
		Onslow et al. (1990)
		Reed & Godden (1977)
	Phase II	Jones et al. (2008)[a]
		Lincoln & Onslow (1997)[b]
		Onslow et al. (1994)
		Rousseau et al. (2007)
		Wilson et al. (2004)
	Phase III	Jones et al. (2005)
		Lewis et al. (2008)
Family-based therapy	Phase I	Millard et al. (2008)*[c]
	Phase II	
	Phase III	
Syllable timed speech	Phase I	Trajkovski et al. (2006)
	Phase II	
	Phase III	

Note. All trials reported an effect except those marked with an asterisk.

[a]Long-term follow-up of the participants in the experimental arm of the Jones et al. (2005) trial.

[b]Long-term follow-up of the participants in the Onslow et al. (1990) and Onslow et al. (1994) trials.

[c]The authors reported that 2 of 6 children showed no clinically significant reduction in stuttering.