Review

The state of the art in non-pharmacological interventions for developmental stuttering. Part 1: a systematic review of effectiveness

Susan Baxter†, Maxine Johnson†, Lindsay Blank†, Anna Cantrell†, Shelagh Brumfitt‡, Pam Enderby† and Elizabeth Goyder†

†School of Health and Related Research, University of Sheffield, Sheffield, UK
‡Department of Human Communication Sciences, University of Sheffield, Sheffield, UK

Abstract

Background: The growing range of available treatment options for people who stutter presents a challenge for clinicians, service managers and commissioners, who need to have access to the best available treatment evidence to guide them in providing the most appropriate interventions. While a number of reviews of interventions for specific populations or a specific type of intervention have been carried out, a broad-based systematic review across all forms of intervention for adults and children was needed to provide evidence to underpin future guidelines, inform the implementation of effective treatments and identify future research priorities.

Aims: To identify and synthesize the published research evidence on the clinical effectiveness of the broad range of non-pharmacological interventions for the management of developmental stuttering.

Methods & Procedures: A systematic review of the literature reporting interventions for developmental stuttering was carried out between August 2013 and April 2014. Searches were not limited by language or location, but were restricted by date to studies published from 1990 onwards. Methods for the identification of relevant studies included electronic database searching, reference list checking, citation searching and hand searching of key journals. Appraisal of study quality was performed using a tool based on established criteria for considering risk of bias. Due to heterogeneity in intervention content and outcomes, a narrative synthesis was completed.

Main Contribution: The review included all available types of intervention and found that most may be of benefit to at least some people who stutter. There was evidence, however, of considerable individual variation in response to these interventions. The review indicated that effects could be maintained following all types of interventions (although this was weakest with regard to feedback and technology interventions).

Conclusions: This review highlights a need for greater consensus with regard to the key outcomes used to evaluate stuttering interventions, and also a need for enhanced understanding of the process whereby interventions effect change. Further analysis of the variation in effectiveness for different individuals or groups is needed in order to identify who may benefit most from which intervention.

Keywords: stuttering, stammering, systematic review, clinical effectiveness.

What this paper adds?
What is already known on the subject?
An increasing number of interventions are available for people who stutter. Evidence has suggested the effectiveness of behavioural programmes in particular for children and comprehensive approaches for adults. However, a systematic review of the literature across all types of interventions and literature was required.

Address correspondence to: Susan Baxter, School of Health and Related Research, University of Sheffield, Regent Court, Regent Street, Sheffield S1 4DA, UK; e-mail: s.k.baxter@sheffield.ac.uk

What this study adds?
The review summarizes evidence from a substantial body of work, and indicates the potential for positive outcomes from a range of interventions. It highlights the need to identify key agreed outcomes that should be used both to evaluate and to compare the effectiveness of interventions, and further investigate individual variation in response.

Background

It has been estimated that approximately 5% of children exhibit developmental stuttering in their early years, with around 75% of them recovering spontaneously (Blomgren 2013). Stuttering is a complex disorder of communication, which may encompass not only surface disruption of fluency but also social and emotional elements. Overt stuttering behaviours that may be apparent to a listener include: repetitions and prolongations of sounds, syllables, words or phrases; or involuntary silent pauses or blocks; or hesitation or pausing before speech. Covert behaviours, which may be undetectable to a listener, include the avoidance of particular words or situations.

Treatments for stuttering (which is more often known as stammering in the UK) have been available for children and adults for many decades. These treatments have encompassed diverse techniques from the use of carbon dioxide, or pharmacological interventions, to those that are behaviourally based. Recent published reviews of stuttering interventions have included: Nye *et al.* (2013), which evaluated the effectiveness of nine studies reporting behavioural interventions for children and adolescents; Blomgren (2013), which provided an overview of selected approaches to interventions with children and adults; Sidavi and Fabus (2010), which summarized the literature on interventions with pre-school and elementary school-age children; Blomgren (2010), which provided an overview of treatment approaches for adults; Lincoln *et al.* (2006), which examined the use of altered auditory feedback; Herder *et al.* (2006), which reports a systematic review and meta-analysis of behavioural stuttering treatments; and Bothe *et al.* (2006a, 2006b), which carried out systematic reviews of behavioural treatment studies and pharmacological interventions published up to 2005.

These reviews have highlighted the increasing size of the literature, and the wide variety of approaches that exist. The evidence underpinning pharmacological interventions is described as 'at best, conflicting' (Bothe *et al.* 2006b), with concerns regarding potential side-effects. Other, non-pharmacological approaches may be categorized using a variety of terms. The most common distinction is between those described as behavioural interventions (which change speech behaviour) versus cognitive and related approaches (Herder *et al.* 2006). Interventions for children may be also divided into either direct approaches which intervene with the child, or indirect approaches which target family interactions. Existing reviews have identified evidence of treatment effectiveness for behavioural interventions across adults and children (Nye *et al.* 2013, Blomgren 2013, Bothe *et al.* 2006a). Treatment involving response-contingent principles in particular has been recommended for children, and the use of combined approaches with adults (Bothe *et al.* 2006a).

While providing valuable insights, many of the existing studies used literature review rather than systematic review methods to summarize the literature, and often considered a particular age range (children or adults), or a particular type of interventions (such as behavioural). The most recent systematic reviews which considered a wider range of populations and interventions examined evidence up to nearly 10 years old (Bothe *et al.* 2006a, 2006b). There was therefore a need for an up-to-date systematic review of evidence, and also with the recognition of an increasing array of interventions beyond behavioural programmes, a requirement for an inclusive review which considered all available options for people who stutter (PWS). Also, in recognition of the importance of service users in their care, for a review to include both quantitative and qualitative literature.

Objectives

This work aimed to carry out a systematic review of evidence on interventions to treat developmental stuttering. The objective was to identify, appraise and synthesize systematically quantitative and qualitative evidence on interventions to treat developmental stuttering across all age groups and thus provide a 'state of the art review' in stuttering therapy (Grant and Booth 2009). This paper reports the findings from a review of intervention studies; a companion paper (Johnson *et al.* 2015) considers evidence from review of the qualitative literature.

Methods

A review protocol was developed prior to beginning the study. The protocol was registered with the PROSPERO database number CRD42013004861.

Identification of studies

A systematic and comprehensive literature search of key health, medical and linguistic databases was undertaken between August 2013 and February 2014 to identify studies for potential inclusion. The search strategy was developed by the information specialist on the team (AC), who undertook electronic searching using iterative methods to create a database of citations. Rather than carrying out a single large database search, iterative methods of searching recognize the value of undertaking several consecutive searches, refining the search terms used based on citations retrieved, and also, for example, using key papers identified to search for other papers which cited these key sources or authors. The initial search involved combining terms for the population (PWS), with both broad and more specific terms for interventions. The search strategy that outlines the terms used and search strings is available from the authors if required.

The first main project search was run on Medline (Ovid) and Psycinfo (Ovid) in August 2013. Following refinements to the search terms, a further iteration of the search was then conducted on a larger range of databases in October–November 2013. These databases encompassed: EMBASE; The Cochrane Library (including The Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Database of Abstracts of Reviews of Effects, Health Technology Assessment Database and NHS Economic Evaluations Database; Science Citation Index; Social Science Citation Index; CINAHL; ASSIA; Linguistics and Language Behavior Abstracts; Sociological Abstracts; and the EPPI Centre. In addition to these searches via electronic databases, topic experts and clinicians in the field were consulted for additional search terms, and for suggestions of additional relevant studies or interventions at regular advisory group meetings, and at a clinician workshop session. Citation searching and screening of other reviews was also undertaken, and we also conducted hand screening of key journals in April 2014 to identify any work published since the main searches had been carried out.

Inclusion and exclusion criteria

Searches were limited by date (1990 to present), as the review was aiming to build on previous work by synthesizing the most up-to-date evidence. While we intended that the review would be predominantly limited to work published in English to ensure that papers were relevant to the UK context, we aimed to include where possible international papers reporting any well-evidenced interventions from developed countries (members of the Organisation for Economic Co-operation and Development (OECD)).

We used the established PICOS criteria (Centre for Reviews and Dissemination (CRD) 2008, Schardt *et al.* 2007) to define inclusion and exclusion criteria (see below for a detailed description of the criteria). Due to the expected volume of the literature we included those study designs considered at lower risk of bias, and which offered greater generalizability. However, we set no quality criteria for inclusion beyond that of study design.

Inclusion criteria

Population

- People who stutter (stammer; PWS) of any age (including those with overt stuttering behaviours such as repetition of syllables or blocking, or those with covert behaviours such as word avoidance).
- People with a developmental stutter (rather than of acquired origin).

Interventions

- Any non-pharmacological intervention delivered in any setting which was of potential benefit to PWS (this could be by either reducing the frequency of occurrence of behaviours, or by aiming to address communication and/or social restrictions).

Comparators

- Studies with any comparator (including an alternative intervention, no intervention or usual practice), also studies with no comparator.

Outcomes

- Any outcome relating to a positive effect on the communication or the social and emotional wellbeing of PWS.

Study design

- Designs which may be termed randomized controlled trials, randomized crossover trials, cluster randomized trials, quasi-experimental studies, cohort studies, before and after/longitudinal studies, case-control studies, studies with more than three participants.

Other inclusion criteria

- Studies from any OECD member country.
- Studies published in English and key studies published in other languages with an English abstract.

Selection of papers for review

Citations retrieved via the searching process were uploaded to a Reference Manager database (Version 12). This database of study titles and abstracts was

```
Citations retrieved                    Not stuttering (n=2341)
and screened
(n=4578)                               Design (n=676)

                                       Related to diagnosis (n=481)

                                       Related to outcomes (n=297)

                    Papers             Duplicate (n=260)
                    rejected at
                    title and          Population (n=119)
                    abstract
                    stage              Background (n=75)
                    (n= 4363)
                                       Review papers (n=59)

                                       Pharmacological (n=38)

                                       Language (n=9)

Total full papers screened             Non OECD country (n=7)
(n=215)
                                       Unable to source (n=4)

                                  Full papers excluded (n=86)

                          Total papers identified via reference
                          lists or hand searching (n=8)

                                  Qualitative papers (n=25)

Total full clinical effectiveness
papers included (n=112)
```
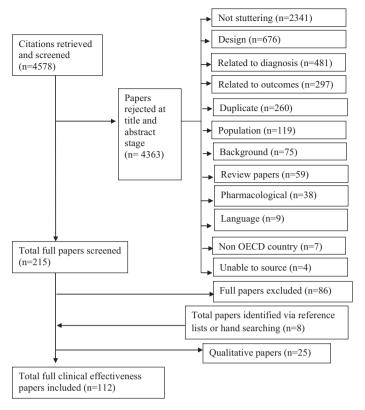
Figure 1. Process of study selection.

independently screened by two reviewers and any difference between reviewers in their inclusion/exclusion decision was resolved by consulting other team members. This screening process entailed the systematic coding of each citation according to its content. The coding included the categorizing of papers falling outside the inclusion criteria (e.g. excluded population, excluded design, excluded intervention) and citations potentially relevant to this review. Full paper copies of all citations coded as potentially relevant were then retrieved for further systematic screening. Papers excluded at this full paper stage and detail regarding the reason for exclusion was recorded. The list of papers excluded at full paper screening stage is available from the authors if required. Figure 1 details the process of selection of papers.

Data extraction

Studies that met our inclusion criteria were read in detail and data extracted. An extraction form was developed using the previous expertise of the review team, to ensure consistency in data retrieved from each study. The completed extraction tables are available from the authors if required. Three members of the research team carried out the data extraction. Data for each individual study were extracted by one reviewer, and in order to ensure rigour, each extraction was checked against the paper by a second member of the team.

Quality appraisal strategy

As the review included a wide range of study designs a comprehensive yet flexible tool was required. Quality assessment was based on the Cochrane criteria for judging risk of bias (Cochrane Collaboration 2011). The assessment tool used within this approach is designed for randomized controlled study designs, therefore we adapted the criteria slightly to make them suitable for observational as well as experimental designs. We also identified aspects within the risk of bias criteria that related particularly to the stuttering literature, such as the use of in-clinic versus real-life situation speech data, and the process of collecting and evaluating the speech sample data (see appendix A).

The summarizing of quality appraisal scoring within and across effectiveness studies is a source of considerable debate in the field of systematic reviewing (Boaz and Ashby 2003, Spencer *et al.* 2003). While some systems total individual elements to provide a score for each study, this practice is now discouraged, as one individual factor may outweigh others in jeopardizing the overall likelihood of bias (Cochrane Collaboration 2011). The risk of bias in a study refers to the potential for the results of a study to differ from the true results. More rigorous studies are more likely to have a lower risk of bias, and yield results that are closer to the truth (Cochrane Collaboration 2011). In order to provide an indication of

stronger or weaker studies across the literature included in this review, we used a system whereby studies were categorized as having either a higher risk of bias, versus a lower risk of bias. The categorization was influenced by an aggregate approach (i.e. how many areas were of concern), but also by considering whether the study contained any particular potential bias that jeopardized the whole study findings. The rating for each study was finalized at a team meeting, with consensus agreement reached. It is important to note that we deliberately used the comparative terms 'higher/lower', and that a 'lower' risk study should not be assumed to be 'low risk'. The rating and any concern regarding quality for each study is detailed in appendix A.

Data analysis and synthesis

The heterogeneity of the included work precluded summarizing the studies via meta-analysis, as will be described below. Findings were reported using narrative synthesis methods, including tabulating characteristics of the included studies, and examining outcomes by type of intervention, outcome measurement, intervention dosage, and length of follow-up.

Results

From a database of 4578 citations, 112 papers met the inclusion criteria for this review. All but two were identified from electronic database searching, with the others being more recent publications which were identified by hand searching. The largest number of papers originated from Australia (39) followed by the United States (26), Canada (11), Germany (9) and the UK (8). Other countries contributing studies included: New Zealand, The Netherlands, Sweden, Turkey, Finland, Ireland, India, Poland and Belgium. Four papers reported data from more than one country. In terms of population, 47 papers described interventions for adults who stutter, 26 school aged children (up to age 11), 15 reported interventions in pre-school children (under age 5), nine mixed age, eight school age and adolescents, and five adolescents only (over age 11). Fifty-one papers reported that interventions were delivered by clinicians, 11 that student clinicians under supervision provided the treatment, the remaining papers were unclear regarding delivery (it was assumed it was the authors).

Study quality

Fourteen papers used comparator group designs. Seven of these were randomized controlled trials, six used quasi-randomization methods (such as consecutive allocation), and one was a controlled before and after study with no randomization of participants. By far the greatest proportion of studies (86) used a before and after design with no comparators, baseline measures pre-intervention and re-measurement after the intervention. One paper used a mixed-method design. Twenty-six papers measured outcomes immediately following the intervention; four at less than 4 weeks follow-up; 23 at 1–6 months; 48 at periods of between 7 and 24 months; and 13 studies had extensive follow-up periods of more than 2 years to more than 5 years. In terms of risk of bias, around two-thirds (77 papers) were rated as being at higher risk of bias and around one-third (35 papers) were judged to be at lower risk of bias.

As detailed above, few of the studies used controlled designs, and of these the allocation process was frequently carried out by pseudo rather than completely randomized procedures (such as consecutive allocation). The areas which tended to distinguish studies rated as having higher potential for bias were firstly, having small samples and a lack of detail regarding the selection and recruitment of the sample. It seemed likely, and was sometimes mentioned, that interventions had been delivered to larger numbers of PWS than were reported. This led to the possibility that those reported differed from those who were not included, with this being a potential significant source of bias. Secondly, a number of studies reported data by individual, rather than pooling findings. Thirdly, a proportion provided limited analysis, for example, reporting only means. Other areas of quality concerns included failing to blind assessors to the time point of data collection, providing limited length of speech data samples, and other concerns regarding the process of data collection, such as participants (or family members) self-selecting speech samples for analysis. In regard to quality in terms of sample size: four studies had large samples of greater than 100 and 12 papers had 50–100. Twenty-three studies included 21–50 individuals, 34 papers had 11–20 participants, and 25 reported 6–10 participants. Thirteen studies had small samples of five or fewer (studies of less than three were excluded), and one did not report the sample size.

Assessment of clinical effectiveness analysed by intervention type

We grouped the papers according to the content of the intervention in order to examine and compare effectiveness. As the literature uses a variety of terms (e.g. 'speak more fluently' versus 'stutter more fluently', 'indirect' versus 'direct', 'speech-restructuring treatment', and 'speech modification therapy') we adopted a classification which endeavours to avoid potential confusion between different authors' use of terminology. The categorization distinguishes seven types of studies: (1) feedback and technology interventions

Table 1. Feedback and technology intervention papers summary

Author, date	Design	Risk of bias	Country	Population
Antipova *et al.*, 2008	Cross-sectional	Higher	New Zealand	Adults, $N = 8$
Armson and Stuart, 1998	Cross-sectional	Higher	Canada	Adults, $N = 12$
Armson *et al.*, 2006	Cross-sectional	Higher	Canada	Adults, $N = 13$
Armson and Kiefte, 2008	Cross-sectional	Higher	Canada	Adults, $N = 31$
Block *et al.*, 2004	Before and after	Higher	Australia	Age 10–16, $N = 12$
Block *et al.*, 1996	Before and after	Higher	Australia	Adults, $N = 18$
Bray and James, 2009	Before and after	Higher	UK	Adults, $N = 5$
Bray and Kehle, 1998	Before and after	Higher	UK	Age 8–13, $N = 4$
Cream *et al.*, 2009	Before and after	Higher	Australia	Adults, $N = 12$
Cream *et al.*, 2010	RCT	Lower	Australia	Adults, $N = 89$
Foundas *et al.*, 2013	Before and after	Higher	USA	Adults, $N = 24$
Gallop and Runyan, 2012	Before and after	Higher	USA	Adults, $N = 11$
Hudock and Kalinowski, 2014	Before and after	Higher	USA	Adults, $N = 9$
O'Donnell *et al.*, 2008	Before and after	Higher	Canada	Adults, $N = 7$
Pollard *et al.*, 2009	Before and after	Lower	USA	Adults, $N = 11$
Ratynska *et al.*, 2012	Cross-sectional	Higher	Poland	Mixed, $N = 335$
Stidham *et al.*, 2006	Before and after	Higher	USA	Adults, $N = 10$
Stuart *et al.*, 2004	Before and after	Higher	USA	Adolescents and adults, $N = 7$
Stuart *et al.*, 2006	Before and after	Higher	USA	Adolescents and adults, $N = 9$
Unger, 2012	Cross-sectional	Higher	Germany	Adults, $N = 30$
Van Borsel, 2003	Before and after	Higher	Belgium	Adults, $N = 9$
Zimmerman *et al.*, 1997	Cross-sectional	Higher	USA	Adults, $N = 9$

Note: RCT, randomized controlled trial.

that aim to change auditory feedback systems (22 papers); (2) cognitive interventions that aim to lead to psychological change (six papers); (3) behavioural modification interventions that aim to change child or parental behaviour, or the behaviour of an adult who stutters (29 papers); (4) speech motor interventions that aim to impact on the mechanisms of speech production such as the respiratory, laryngeal or articulatory systems (18 papers); (5) speech motor combined with cognitive interventions (18 papers); (6) multiple component interventions (11 papers); and (7) studies which compared interventions to each other (eight papers).

Feedback and technology interventions

Twenty-two papers described the effectiveness of a range of technologies aiming to reduce the frequency or severity of stuttering in speech (table 1). The greatest proportion of the technologies described were devices which alter the way that a PWS hears their own speech (altered auditory feedback—AAF). This was effected by changing the frequency (frequency altered feedback—FAF), and/or by introducing a delay before the speech is heard (delayed auditory feedback—DAF). All but two of the papers in this typology (Cream *et al.* 2009, Pollard *et al.* 2009) were rated as being at higher risk of bias.

Use of the SpeechEasy device was reported in six papers (Armson *et al.* 2006, Armson and Kiefte 2008, Pollard *et al.* 2009, Foundas *et al.* 2013, O'Donnell

et al. 2008, Gallop and Runyan 2012). No papers used a control group design, and five of the six were rated as being at potential higher risk of bias. All studies reported some degree of reduction in stuttering rate when the device was in place (compared with no device) including effect sizes of 0.10 and 0.72 (Armson *et al.* 2006, Armson and Kiefte 2008), also self-reported stuttering rate improvement (effect size 0.65; Armson and Kiefte 2008). No adverse effect on speech rate was reported in any of these studies. Longer-term follow-up data from one study indicated some persistence of effect. Gallop and Runyan (2012), however, highlighted that considerable individual variation was apparent in outcomes, with eight of 11 users in one study, and five of seven in another study (Gallop and Runyan 2012, Pollard *et al.* 2009) reporting improved fluency at follow-up, whereas others disliked using the device, or stuttered more with it in place.

Six papers reported the use of other feedback devices combining DAF and FAF including the Pocket Speech Lab, SmallTalk, Fluency Enhancer, Digital Speech Aid, Edinburgh Masker, and a self-contained in-ear fluency device. All these studies were considered to be at higher risk of bias. All findings indicated a positive effect on fluency (reduction in the percentage of words stuttered or syllables stuttered, or improved scores on the Perceptions of Stuttering Inventory—PSI; Woolf 1967). The reported improvements included: a reduction of 2.1% (Block *et al.* 1996); 3% (Antipova *et al.* 2008); 55–60% (Zimmerman *et al.* 1997); 65% (Stuart *et al.* 2004); and 75% (Stuart *et al.* 2006) in words or

Table 2. Cognitive interventions summary

Author, date	Design	Risk of bias	Country	Population
Amster and Klein, 2008	Before and after	Higher	USA	Adults, $N = 8$
De Veer *et al.*, 2009	RCT	Higher	Netherlands	Adults, $N = 37$
Kaya and Alladin, 2012	Before and after	Higher	Turkey	Adults, $N = 59$
Kaya, 2011	Before and after	Higher	Turkey	Adults, $N = 93$
Leahy and Collins, 1991	Before and after	Higher	Ireland	Adults, $N = 5$
Reddy *et al.*, 2010	Before and after	Higher	India	Adults, $N = 5$

syllables stuttered with device versus no device. Also reported were odds ratio of dysfluency with device of 0.18 compared with an odds ratio of 0.58 for no device (Ratynska *et al.* 2012). Individual variation in response was, however, highlighted in these studies. For example, Ratynska *et al.* (2012) noted that while moderate or considerable improvement was found for 84.5% of participants, deterioration or lack of improvement was found for 15.5%. Another paper reported that some individuals increased stuttering on some tasks with the device in place (Block *et al.* 1996).

Other types of technology evaluated in the literature were: bone conduction stimulation; electromyography (EMG); and video self-modelling (VSM). Bone conduction stimulation (Wilson *et al.* 2004) seemed to have a temporary effect (up to 2 weeks), however, required an uncomfortable and obtrusive headband to be worn. The only technology trialled with children was EMG feedback. The two studies evaluating this type of intervention reported improvements (a reduction to less than 1% stuttered syllables by Craig *et al.* 1996; and a mean 37% reduction by Block *et al.* 2004), however, there were some concerns regarding a negative impact on speech rate. The three studies investigating VSM found a reduction in stuttered words or syllables (Bray and Kehle 1998, Cream *et al.* 2009) and a positive impact on self-reported severity (Cream *et al.* 2009). However, in another study VSM was found to have limited impact on fluency when added to a maintenance programme, compared with standard maintenance (Cream *et al.* 2010).

Cognitive interventions

This type of intervention may be used alone, or to support, optimize or prepare for other interventions. The anticipated outcomes may be direct speech gains, psychological well-being gains which lead to improved speech, or alternatively gains which do not aim to change the frequency or severity of the stutter, but instead relate to living successfully with stuttering. As may be expected, this body of work used a wider range of outcome measures than those evaluating technological devices. Six papers were identified within this intervention (table 2), with other work combining cognitive approaches with other elements reported in a following section.

Two papers evaluated cognitive behaviour therapy (CBT), and while one provided only limited mostly descriptive data (Reddy *et al.* 2010) the other (Amster and Klein 2008) provided evidence of positive outcomes in terms of perfectionism, communication attitudes ($p = 0.017$) and speech fluency (ES = 0.51). This intervention included speech motor elements in the final few sessions. Leahy and Collins (1991) reported that Personal Construct Psychology sessions could have a positive effect on Stuttering Severity Index (SSI; Riley 1994) scores (although it appeared that participants had also received other concurrent therapy). The use of hypnosis was evaluated in two papers (Kaya 2011, Kaya and Alladin 2012) with a significant positive effect on ratings of fluency ($p < 0.000$) reported. The rating scale used in this study, however, had considerable limitations.

The Mindfulness-Based Stress Reduction program was evaluated by De Veer *et al.* (2009). Although this was the only study in the group that used a controlled design, the recruitment and allocation process and self-report measures resulted in a rating of higher potential for bias. Average effect sizes were found for self-efficacy beliefs, coping and attitude towards speech situations ($d = 0.55$, 0.62 and 0.48, respectively). Effect sizes were large for stress, anxiety and locus of control ($d = 1.16$, 1.07 and 0.76 respectively). There was some maintenance of these positive outcomes at 4 weeks.

Behaviour modification

The greatest number of papers identified which related to a single intervention was the Lidcombe Program (LP). This intervention for children is based on operant conditioning principles with the content focusing on training parents to provide feedback (verbal contingencies) for stuttered speech and stutter-free speech. Twenty-two papers considered aspects of the program, including effectiveness in the short term and longer-term, predictors of treatment time, predictors of responsiveness, applicability in different countries, and components of intervention delivery such as tele-health. This intervention group contained 12 papers assessed as being at lower risk of bias (table 3). It is important to note that when considering interventions with young

Table 3. Papers relating to the Lidcombe Program summary

Author, date	Design	Risk of bias	Country	Population
Bonelli *et al.*, 2000	Before and after	Higher	Australia	Pre-school, $N = 9$
Femrell *et al.*, 2012	Before and after	Higher	Sweden	Pre-school, $N = 10$
Harris *et al.*, 2002	Quasi-RCT	Lower	Australia	Children—age unclear, $N = 23$
Harrison *et al.*, 2004	Before and after	Lower	Australia	Pre-school, $N = 38$
Jones *et al.*, 2000	Before and after	Lower	Australia	Pre-school, $N = 216$
Jones *et al.*, 2005	RCT	Lower.	New Zealand	Pre-school, $N = 54$
Jones *et al.*, 2008	RCT	Lower	Australia/New Zealand/USA	School age, $N = 28$
Kingston *et al.*, 2003	Before and after	Higher	UK	Pre-school, $N = 78$
Koushik *et al.*, 2009	Before and after	Lower	Canada	School age, $N = 11$
Koushik *et al.*, 2011	Cross-sectional (retrospective case note analysis)	Higher	USA	Pre-school, $N = 134$
Lattermann *et al.*, 2008	Quasi-RCT	Lower	Germany	Pre-school, $N = 45$
Lewis *et al.*, 2008	RCT	Lower	Australia	Pre-school, $N = 18$
Lincoln and Onslow, 1997	Cross-sectional (follow-up data)	Higher	Australia	School age, $N = 43$
Lincoln *et al.*, 1996	Before and after	Higher	Australia	School age, $N = 11$
Miller and Guitar, 2009	Cross-sectional (long-term outcomes data)	Lower	USA	Pre-school, $N = 15$
O'Brian *et al.*, 2013	Before and after	Lower	Australia	Pre-school, $N = 57$
Onslow *et al.*, 1994	Controlled before and after	Higher	Australia	Pre-school, $N = 11$
Onslow *et al.*, 1990	Before and after	Higher	Australia	Pre-school, $N = 4$
Onslow *et al.*, 2002	Cross-sectional	Higher	Australia	School age, $N = 8$
Rousseau *et al.*, 2007	Before and after	Lower	Australia	Children—unspecified age, $N = 29$
Wilson *et al.*, 2004	Before and after	Higher	Australia	Pre-school and school age, $N = 5$
Woods *et al.*, 2002	Before and after	Lower	Australia	Pre-school and school age, $N = 8$

children interventions need to demonstrate not only evidence of effectiveness but change beyond a level of spontaneous recovery. There is currently uncertainty regarding the precise figure, however. Yairi and Ambrose (1992) estimated recovery rate from transient stuttering to be up to 80% by adolescence/adulthood.

Of the 11 papers focusing primarily on clinical effectiveness of the LP, all studies found positive effects on the percentage of syllables stuttered (%SS), and indicated benefits continuing post-intervention. The studies using controlled designs indicated that, while children improved without receiving the programme, the intervention had superior outcomes to spontaneous recovery for most participants. Harris *et al.* (2002), for example, found a significant mean reduction in %SS of 39% ($p < 0.001$) pre- to post-intervention in the nine children in the LP intervention group. This compared with a reduction of 16% for nine of the children who had not received the intervention, and an increase of 6–54%SS in four other children in the control group. Jones *et al.* (2005, 2008) also used a controlled trial design with 16 of 19 participants who had completed the intervention and could be contacted, assessed as having zero to 1.1%SS at 5-year follow-up, and three exhibited some relapse. Five of the eight participants who could be contacted in the control group were reported to have recovered spontaneously. A study in Germany (Lattermann *et al.* 2008) echoed these positive findings (reduction of 6.8%SS intervention versus 3.6% control), and

confirmed that the reduction in %SS was not at the expense of a reduction in rate of speech. Other papers which investigated delivery issues and predictors of outcome, indicated that baseline severity and time since onset may contribute to variation in response, and that specialist training for clinicians was important (O'Brian *et al.* 2013, Miller and Guitar 2009, Kingston *et al.* 2003, Jones *et al.* 2000). Papers by Woods *et al.* (2002), Onslow *et al.* (2002) and Bonelli *et al.* (2000) indicated no adverse effect of the programme on the child or parents.

Other interventions which we categorized as behaviour modification, focused on changing behaviours within the family, predominantly parent behaviour and parent–child interaction. Four papers were identified which evaluated these interventions in children, and all were rated as being at higher risk of bias (table 4). Two additional papers in this category evaluated behaviour modification programmes with adults.

Papers in this group all reported positive outcomes following interventions. Four of six children in one study, and four of six in a second study, reduced stuttering severity on a rating scale following Parent–Child Interaction therapy, with the effect persisting at 1 year follow-up (Millard *et al.* 2008, 2009). A family-focused treatment resulted in a significant reduction in dysfluencies ($p < 0.001$), and following the Rustin family-oriented program (Mallard 1998) 23 of the 28 families (82%) did not require any further intervention.

Table 4. Non-Lidcombe behaviour modification interventions summary

Author, date	Design	Risk of bias	Country	Population
Franklin *et al.*, 2008	Quasi-RCT	Higher	Australia	Adults, $N = 60$
Hewat *et al.*, 2006	Quasi-RCT	Lower	Australia	Adolescents and adults, $N = 30$
Lutz, 2009	Before and after	Higher	Germany	Parents, $N = 11$
Mallard, 1998	Before and after	Higher	USA	School age, $N = 28$
Millard *et al.*, 2008	Before and after	Higher	UK	Pre-school, $N = 6$
Millard *et al.*, 2009	Before and after	Higher	UK	Pre-school, $N = 10$
Yaruss *et al.*, 2006	Before and after	Higher	USA	Pre-school, $N = 17$

Table 5. Speech motor interventions summary

Author, date	Design	Risk of bias	Country	Population
Andrews *et al.*, 2012	Before and after	Higher	Australia	School age, $N = 10$
Block *et al.*, 2005	Before and after	Lower	Australia	Adults, $N = 80$
Block *et al.*, 2006	Before and after	Lower	Australia	Adults, $N = 80$
Carey *et al.*, 2010	Before and after	Lower	Australia	Adults, $N = 40$
Cocomazzo *et al.*, 2012	Before and after	Lower	Australia	Adults, $N = 12$
Druce *et al.*, 1997	Before and after	Lower	Australia	School age, $N = 15$
Franken *et al.*, 1992	Before and after	Higher	Netherlands	Adults, $N = 32$
Ingham *et al.*, 2013	Before and after	Higher	USA	Adults, $N = 30$
Ingham *et al.*, 2001	Before and after	Higher	USA	Adults, $N = 5$
Iverach *et al.*, 2009	Before and after	Lower	Australia	Adults, $N = 64$
O'Brian *et al.*, 2003	Before and after	Higher	Australia	Adults, $N = 30$
O'Brian *et al.*, 2008	Before and after	Higher	Australia	Adults, $N = 10$
Onslow *et al.*, 1992	Before and after	Higher	Australia	Adolescents and adults, $N = 14$
Onslow *et al.*, 1996	Before and after	Higher	Australia	Adults, $N = 18$
Trajkovski *et al.*, 2011	Before and after	Lower	Australia	Pre-school, $N = 17$
Von Gudenberg, 2006	Before and after	Higher	Germany	Unclear, $N =$ unclear
Von Gudenberg *et al.*, 2006	Before and after	Higher	Germany	School age, $N = 32$
Yairi and Ambrose, 1992	Before and after	Higher	USA	Pre-school, $N = 27$

An additional intervention that we categorized as behaviour modification is used with adults who stutter. Self-imposed time-out treatment is, like the LP, based on an operant conditioning approach. This intervention was evaluated in two papers, one rated as lower (Hewat *et al.* 2006) and one as higher risk of bias (Franklin *et al.* 2008). Both papers reported a mean reduction in stuttering (of 53.6%SS and from 5.8%SS to 3.9%SS), although, while there was no evidence of a negative impact on speech rate, there was considerable individual diversity in response.

Speech motor interventions

Eighteen papers evaluated interventions which we classified in our typology as speech motor interventions. The content of these interventions is focused on the mechanisms of speech production (breathing, vocal fold vibration, articulation of sounds) with reduction in the severity or frequency of stuttering achieved by altering speech motor patterns. While these interventions were given various labels, the largest group were described as focusing on teaching prolonged speech (PS). In this group four (mostly older papers) were graded as being at higher risk of bias, and three as being at lower risk of bias (table 5).

The papers all reported improvement in %SS following PS intervention and at up to 5-year follow-up. O'Brian *et al.* (2003), for example, found pre-treatment mean 7.9%SS reduced to 0.4%SS at 12 months. Block *et al.* (2006) reported a pre-treatment mean %SS of 4.9%SS, reduced to a mean of 0.9%SS immediately post-treatment, and 2.6%SS at 1 year (significant difference from baseline $p = 0.04$), and at 3.5–5 years, during a surprise telephone call, the mean %SS was 1.6. Evaluation of a Smooth Speech intervention (Block *et al.* 2005) similarly found the reduction in %SS following treatment was statistically significant (pre-treatment 5.4%SS and post-treatment 1.8%SS, large ES 0.86) and at 3.5–5-year FU the mean stuttering rate was 1.6%SS. Assessment of mean naturalness scores in these studies indicated that post-intervention there was no significant impact on naturalness (Onslow *et al.* 1992, 1996), and speech rate also did not appear to be adversely affected. (O'Brian *et al.* 2013). Four papers evaluating speech motor interventions with children also reported positive outcomes, including an effect size of $d = 0.096$ (Von Gudenberg 2006), and a reduction in mean %SS from 7.6 to 1.75%SS ($p = 0.0015$) (Druce *et al.* 1997).

The final papers in this group (Andrews *et al.* 2012, Trajkovski *et al.* 2011) examined interventions for pre-school and school aged children termed Syllable Timed

Table 6. Speech motor plus cognitive interventions summary

Author, date	Design	Risk of bias	Country	Population
Baumeister *et al.*, 2003	Before and after	Higher	Germany	School age and adolescents, $N = 37$
Beilby *et al.*, 2012	Before and after	Lower	Australia	Adults, $N = 20$
Berkowitz *et al.*, 1994	Before and after	Higher	USA	School age, $N = 8$
Blomgren *et al.*, 2005	Before and after	Lower	USA	Adults, $N = 19$
Boberg and Kully, 1994	Before and after	Higher	USA	Adolescents and adults, $N = 49$
Huinck *et al.*, 2006	Before and after	Lower	Netherlands	Adults, $N = 25$
Irani *et al.*, 2012	Mixed method	Higher	USA	Adults, $N = 7$
Laiho and Klippi, 2007	Before and after	Higher	Finland	School age, $N = 21$
Langevin and Boberg, 1993	Before and after	Higher	Canada	Adults, $N = 10$
Langevin and Boberg, 1996	Before and after	Lower	Canada	Adults, $N = 4$
Langevin *et al.*, 2006	Before and after	Lower	Canada and Netherlands	Adults, $N = 25$
Langevin *et al.*, 2010	Before and after	Lower	Canada	Adults, $N = 17$
Lawson *et al.*, 1993	Before and after	Higher	UK	Adolescents, $N = 15$ and 19
Nilsen and Ramberg, 1999	Before and after	Higher	Sweden	Adolescents, $N = 13$
Rosenberger, 2007	Before and after	Higher	Germany	School-age children and adolescents, $N = 19$ and 15
Smits-Bandstra and Yovetitch, 2003	Before and after	Higher	Canada	School age, $N = 3$
Stewart, 1996	Before and after	Higher	UK	Adults, $N = 12$
Ward, 1992	Before and after	Higher	UK	Unclear, $N = 4$

Speech (STS), and reported mean stuttering reductions of 96%SS (6%SS to 0.2%SS) at 12 months follow-up (ES = 1.8), and 14.4%SS to 6.7% (ES = 0.7). Data on self-reported severity, situation avoidance and quality of life confirmed these positive outcomes, however the authors noted considerable individual variation in response to the intervention.

Speech motor combined with cognitive elements

As mentioned above, cognitive interventions may be used as an intervention type in isolation, or alternatively may form part of a programme. Eighteen included papers reported interventions which combined speech motor therapy with elements of cognitive interventions. A third of these papers combining cognitive with speech motor elements were judged to be at lower risk of bias (table 6).

All papers in this group reported a positive impact on clients following intervention (including a 55–99% reduction; a 38% reduction; from 15.3%SS to 2.4%SS; from 19.59%SS to 1.29%SS; from 10.2 to 3.3%SS; from 6.42 to 1.39; ES *d* = 2.07, 0.69, 1.29 and 1.12 respectively). Some papers used measures other than reduction in syllables stuttered which also indicated benefits for participants following treatment. Blomgren *et al.* (2005) found improvements on both the Avoidance and Expectancy subscales of the PSI ($p < 0.001$). Laiho and Klippi (2007) and Lawson *et al.* (1993) echoed the positive effect on avoidance. Blomgren *et al.* (2005) also noted change on two specific affective functioning measures—the Psychic and Somatic Anxiety subscales of the Multi-Component Anxiety Inventory (MCAI-IV) with a significant difference of $p = 0.078$ and $p = 0.036$ respectively. Rosenberger *et al.* (2007) similarly

recorded a positive effect on anxiety ($p < 0.025$). Irani *et al.* (2012) outline improvement in S-24 (Modified Erikson Scale of Communication Attitudes; Andrews and Cutler 1974) scores with an ES of 1.19, and Locus of Control of Behavior Scale (Craig *et al.* 1984) improvement with an ES of 0.75. In another study (Beilby *et al.* 2012) the psychological measures used also improved significantly post-intervention and at FU ($p < 0.001$).

Stewart (1996) noted an interesting relationship between the two elements of these programmes: that specific attitude change sessions did not seem to result in significant changes to attitude measures, however change was apparent in most of the attitude measures following the technique sessions. While two papers in this group noted some fading of effect, the authors of these studies described no adverse effect on speech fluency.

Multiple elements

Eleven papers described interventions which included multiple components across our typology of interventions, or were papers which evaluated a range of interventions (table 7). These treatment programmes included combinations of elements including: EMG feedback; smooth speech; relaxation; airflow training; cognitive aspects; self-management components; social support; and particular elements to prevent relapse, and all reported positive outcomes. Two papers (Hancock and Craig 2002, Hancock *et al.* 2002) highlighted the value of a multiple element re-treatment programme for adolescents, although considerable individual differences were noted. Other papers in this group reported: reduction in number of stutters from 5.34% to 0.18%; reduction of stuttering to less than 3%SS for

Table 7. Papers reporting multiple component interventions summary

Author, date	Design	Risk of bias	Country	Population
Allen, 2011	Cross-sectional	Higher	UK	Adults, $N = 16$
Blood, 1995	Before and after	Higher	USA	Adults, $N = 4$
Craig *et al.*, 2002	Before and after	Higher	Australia	Adolescents, $N = 6$
Elliott *et al.*, 1998	Before and after	Higher	USA	School age, $N = 5$
Gagnon and Ladouceur, 1992	Before and after	Higher	Canada	School age, $N = 4$ and 4
Hancock and Craig, 2002	Before and after	Lower	Australia	Adolescents, $N = 12$
Hasbrouck, 1992	Before and after	Higher	USA	Adults, $N = 117$
Pape-Neumann, 2004	Before and after	Higher	Germany	Adults, $N = 100$
Sicotte *et al.*, 2003	Before and after	Higher	Canada	School age, $N = 6$
Wagaman *et al.*, 1993	Before and after	Higher	USA	School age, $N = 8$
Wagaman *et al.*, 1995	Before and after	Higher	USA	Adolescents, $N = 7$

Table 8. Papers comparing interventions summary

Author, date	Design	Risk of bias	Country	Population
Craig *et al.*, 1996	Quasi-RCT	Lower	Australia	Children and adolescents, $N = 97$
Franken *et al.*, 2005	Before and after	Lower	Netherlands	Pre-school, $N = 23$
Hancock *et al.*, 1998	Cross-sectional (further analysis of RCT data)	Lower	Australia	Children and adolescents, $N = 77$
Hancock and Craig, 1998	Cross-sectional (further analysis of RCT data)	Lower	Australia	Adolescents, $N = 97$
Menzies *et al.*, 2008	RCT	Lower	Australia	Adults, $N = 30$
Riley and Ingham, 2000	Before and after	Higher	USA	School age, $N = 12$
Ryan and Van Kirk, 1995	Before and after	Lower	USA	School age and adolescents, $N = 24$
Wille, 1999	Before and after	Higher	Germany	School age and adolescents, $N = 14$

all participants; positive changes in thinking and feeling scales. There was evidence of long-term impact up to 5 years (Wagaman *et al.* 1995).

One paper (Pape-Neumann 2004) highlighted the range of programmes used by clinicians. For any intervention delivered to children data from this study indicated an intervention effect size of 0.63. For adolescents and adults the effect size across any intervention was 0.77. A positive impact on avoidance of communication, attitude towards communication, self-judgement of stammering in social situations, and everyday life was also recorded across the therapies.

Papers comparing interventions

We identified eight papers which compared interventions with one another (rather than having no comparator, or comparing an intervention with no intervention). These papers were generally of reasonable quality, with only two considered to be at higher risk of bias. Only one compared interventions delivered to adults (table 8).

The LP was compared with treatment based on the Demands and Capacities Model (DCM) in pre-school children (Franken *et al.* 2005). For the LP the mean stuttering frequency decreased from 7.2% to 3.7%, and the DCM treatment mean decreased from 7.9% to 3.1%. Stuttering severity ratings significantly reduced for both interventions, with no significant difference between them ($p > 0.10$). Speech motor intervention alone was compared with speech motor combined with

CBT in adults who stutter (Menzies *et al.* 2008). The authors found no difference in effect on %SS between the two groups although the group who received additional CBT showed greater improvement on measures of anxiety and avoidance. Data are limited in a paper comparing bio-resonance therapy to 'standard therapy' resulting in it not being possible to conclude which was more effective (Wille 1999). Another comparison study found that both the method of gradually increasing length and complexity of utterances (GILCU) and DAF, led to significant reductions in %SS ($p > 0.01$), however DAF was found to be slightly superior than GILCU in terms of generalizability (Ryan and Van Kirk 1995). EMG was compared with both intensive smooth speech and home based speech programmes for children and adolescents by Craig *et al.* 1996. The three treatments were found to be equally effective ($p = 0$). Only one paper of the eight in this group found a significant difference between the interventions which were compared. Riley and Ingham (2000) found that while both were effective, an extended length of utterance (ELU) intervention had a significantly greater effect ($p = 0.04$) than speech motor training (SMT).

Effectiveness across the typologies

The most significant challenge in comparing clinical effectiveness between different studies and types of therapy, is the range of outcome measures used. The outcomes most often evaluated were those relating to the overt frequency or severity of stuttering.

However, within this category authors used a variety of specific measures including: number of words stuttered; percentage of words stuttered; number of syllables stuttered; percentage of syllables stuttered; number of stuttering events; number of dysfluencies; and various ratings of stuttering severity. The measurement could be carried out by the research team in person, by telephone, or via recorded speech samples sent by participants. Alternatively, assessment could be carried by an independent listener from audio or video recordings, or by a clinician. In a minority of studies PWS provided a self-evaluation of stuttering frequency. In addition to these frequency and severity measures some studies also used standardized and non-standardized assessments measuring perceived self-efficacy/control/esteem; anxiety/stress/depression level; self-perceptions of or attitude to speech/stutter; perception of self/others as being a stutterer; avoidance of words/situations; parent verbal interaction; rate of speech; and perceived naturalness or fluency. A few studies included evaluation of participant satisfaction with the intervention.

In addition to a wide variety of outcome measure, the literature used three main strategies to evaluate effectiveness in terms of frequency or severity of stuttering: firstly, by comparing percentage change pre to post; secondly, by reporting level of frequency of stuttering at baseline, and again post-intervention then using statistical means to examine the difference; and thirdly, by descriptively comparing the severity level or need for further intervention pre–post.

Much of the literature reported effectiveness via the first of these methods, the percentage of reduction in stuttered speech pre- and post-intervention (most often change in the percentage of syllables stuttered). These 'degree of change' measures however have a significant limitation when comparing studies, as baseline stuttering severity will influence how substantial any positive change can be. An examination of baseline levels of stuttering amongst participants across studies revealed a high level of variation in the fluency of participants prior to the intervention both between studies, and importantly within a study. For example, Lincoln *et al.* (1996) reported a baseline mean amongst participants ranging from 5%SS to 18.9%SS, whereas Langevin and Boberg (1993) found a baseline stuttering rate ranging from 3.6%SS to 9.4%SS and participants in the Rousseau *et al.* (2007) study appeared to have a low baseline of mean 3%SS. Many studies highlighted that there was considerable individual variation in outcome (e.g. Andrews *et al.* 2012, Block *et al.* 2006, Foundas *et al.* 2013), with some linking this to baseline level of severity (e.g. O'Brian *et al.* 2013). The link between baseline severity and outcome, suggests that comparison of effectiveness between different studies with dif-

ferent participant characteristics needs to be treated with caution.

While these concerns regarding comparison should be fully considered, we were interested to descriptively examine and compare studies using this form of reporting of change. Within the 'feedback and technology' group authors reported percentage reductions in syllables stuttered of between 3% and 87%. In the 'behaviour modification' typology, percentage reductions in syllables stuttered ranged from 69% to 97% for the LP, and 53% for the one paper reporting this measure using other interventions. In the 'speech motor' group %SS reductions were reported varying from 22% to 96%. For 'speech motor plus cognitive' the one paper using this measure found a 22% reduction in %SS. Amongst the 'multiple components' group, reductions of 52% and 89% were described, and a 36.5% and 63.5% improvement in a comparison paper evaluating ELU and SMT interventions was described. There was evidence therefore that large reductions in stuttering frequency could result from interventions across the typologies.

In addition to the requirement to consider baseline stuttering levels when evaluating these papers, it is also important to consider the impact of any change for the person who stutters. This requires consideration of not only change but clinical (or personal) impact of the improvement. It was proposed in one included study that in order to be clinically significant, an intervention should result in a 50% reduction in stuttering (Reddy *et al.* 2010). While this threshold is not universally accepted as a valid outcome (see Discussion section) examining the set of papers reporting percentage change in syllables stuttered with this criterion reveals that six of the eight 'feedback and technology' group, all five of the Lidcombe papers, one non-LP behaviour modification paper, five of the six 'speech motor' papers and both of the multiple component interventions reached this level, again indicating that a range of intervention approaches identified could result in clinically significant improvement.

The second method for evaluating change was to report effectiveness via statistical significance (*p* values or effect sizes), rather than percentage change. In the 'feedback and technology' group an effect size of 0.14 was reported for stuttering (during monologue only) in one paper, and an effect size of 1.1 for reduction of stuttering frequency in another, and statistically significant changes were reported in nine papers. In the 'cognitive' group an effect size of 0.74 for reduction in stuttering was found in one study, and three papers reported statistically significant differences. In the behaviour modification group effect sizes of 2.3, 2.3 and 2.9 are described, and eight papers provided evidence of statistically significant effects. Speech motor and speech motor plus cognitive approach papers reported effect

sizes of 0.7, 0.86, 0.88, 0.96, 1.29, 1.12, 1.8, 6.86 and 14.96, and seven reported statistically significant effects. One multiple component paper reported an effect size of 0.63, and three comparison papers described significant effects for each of the interventions they evaluated. This set of findings thus supports the conclusion that a range of interventions may be effective for PWS. As with the percentage change evaluations described above statistical significance may differ from clinical (or personal) significance and mask individual variation in outcomes with some studies reporting large confidence intervals. Statistical significance is also heavily dependent upon sample size, as very small effects can be statistically significant with a large sample, while relatively large effects may not be statistically significant with a smaller sample. Using effect sizes accounts for differences in sample size thereby giving an appropriate weighting to studies results.

The third approach to evaluating stuttering frequency or severity outcomes considered whether desired levels of fluency had been achieved, used scaled data to judge severity, or reported whether further intervention was required. While there is considerable debate regarding what is a 'good outcome' in terms of the level of stuttering following an intervention, many studies use a 3%SS or less level as being an acceptable degree of dysfluency, and thus this may be a target for interventions to achieve (Gagnon and Ladouceur 1992). Four papers in the 'cognitive' group used severity scales to evaluate difference (two a scale developed by the author, and two the SSI), these studies found positive outcomes. The papers reporting the LP often included data from severity rating scales, and the programme uses threshold levels of %SS in order for participants to move through the intervention stages. Four non-LP 'behaviour modification' papers reported positive outcomes in terms of parent report, stuttering severity or need for further therapy post-intervention. Five 'speech motor' papers reported reduced levels of stuttering post-therapy (to 0.9%SS, 1.6%, 1%, 0.4%, near 0%), three 'speech motor plus cognitive' (to 1.29%, 0.53%, 0.1–3.8%), six multiple component papers (to less than 3% in four papers, less than 2% in two). These papers further confirm that using this approach to measuring effectiveness there is evidence of positive outcomes for PWS across a range of intervention approaches.

While overt stuttering frequency or severity measures were the most frequently reported outcome, a smaller number of papers considered wider effects on the person who stutters or self-rated perceptions of stuttering. One feedback and technology paper (Pollard *et al.* 2009) used PSI scores. This paper found however, that the significant effect of the technology immediately post-fitting was not maintained at follow-up. The 'cognitive' interventions group (as may be expected) tended to use a wider range of measures to evaluate efficacy, indicating that the intervention could impact on not only stuttered speech but also self-perceptions and attitudes. De Veer *et al.* (2009), for example, reported large effect sizes on anxiety and locus of control. In the 'speech motor and cognitive category' Lawson *et al.* (1993) found change in PSI scores, with reduction in avoidance the greatest area of change.

Dose–response outcomes

We endeavoured to explore whether the number of hours of intervention could be linked to outcomes for PWS. The heterogeneity in measures used, variation in time points assessed and limited detail in some papers made this type of analysis problematic, however in order to explore this potential relationship we examined those papers which included statistical analysis (*p* values or effect sizes). Not only did different measures preclude drawing robust conclusions regarding a relationship, but also there was a substantial body of the literature which reported that intervention hours varied between individuals receiving the same intervention. From our limited examination of studies there was no clear evidence that increasing contact hours for all participants led to more positive outcomes, and there was little discussion in the literature regarding how the contact hours had been determined.

Discussion

This review of intervention effectiveness found a substantial body of work (112 papers), which we divided into a typology of seven categories. Across the set of papers, the overall conclusion is that a diverse range of intervention types delivered by clinicians have some evidence of effectiveness underpinning their use for PWS, with most intervention studies able to demonstrate a positive effect for at least some participants. The review findings echo much earlier work by Bloodstein (1987) who similarly concluded that a significant proportion of participants will gain from interventions, no matter what type, and also the more recent Herder *et al.* (2006) review. While the range of interventions and volume of research has grown considerably since this work, it is perhaps disappointing that the evidence remains unclear regarding who will gain most from which programme.

The current review presents a more encouraging message regarding the diversity of effective interventions than a previous review (Bothe *et al.* 2006a), which used more stringent inclusion criteria for measurement of outcomes. Both examinations of the evidence however provide generally positive results. Compared with the Nye *et al.* (2013) meta-analysis of behavioural

interventions for children, and Sidavi and Fabus's (2010) review, the present review finds support for not only the LP, but a range of interventions, although evidence for the LP is underpinned by a substantial body of evidence and a funded programme of evaluation. This review highlights that while researchers continue to build evidence underpinning individual programmes, the comparison of interventions with each other remains problematic. In order to advance understanding there is a need for future intervention studies to pay greater attention to variables such as participant and clinician characteristics, time since onset, session time: number of sessions per week, minutes per session in order to make comparative investigations feasible.

It is important to note that the literature has considerable variation in quality, with around three-quarters of the studies rated as being at higher risk of bias. Our quality rating of higher/lower risk was a comparative one, with only a very small proportion of the literature using the highest quality controlled designs and as a whole therefore there is little that could be considered to be at low risk of bias. In the lower quality studies in particular, the potential for participants reported to differ from those not recruited and/or reported cannot be ruled out. The higher quality studies however did not report contradictory findings to the poorer ones, and also on a positive note, the body of work examined contains a sizeable number of studies (25) with lengthy follow-up periods. Twenty-five papers had a follow-up period of 2 years or more providing evidence that while some fading of effect was likely, that positive outcomes could persist in the longer-term. The only group of interventions where effects seemed of shorter term duration, was the feedback and technology group which may offer PWS a more immediate gain in fluency to be used in particular situations of difficulty (such as talking on the telephone). In a review specifically of altered auditory feedback, Lincoln *et al.* (2006) highlighted a lack of evidence of effectiveness in everyday situations. Many of the papers we included have been published since that review, with the literature seemingly beginning to address this issue.

In the generally positive reporting of study findings however, there was, in many cases a sizeable number of participants who did not achieve benefit or achieved poorer outcomes than the rest. Differing responses have also been noted in other reviews (Blomgren 2010, Lincoln *et al.* 2006). The individual variability in response was substantial across the set of the literature we examined, with evidence that any intervention would not be successful for all who received it. The current review therefore highlights the limitations of a 'one size fits all' approach to offering treatment. Instead the findings suggest a need for a range of options to be available to a person who stutters.

The comparison of stuttering interventions with each other is hampered by variation in systems of measurement, and variation in intervention contact hours. There is little available research which compares the effectiveness of different interventions and thus a very limited pool of evidence for clinicians and PWS to draw on in selecting an optimal intervention. We were unable to demonstrate any clear dose–response relationship across intervention typologies, meaning that currently interventions with many hours of contact did not seem to offer substantially different outcomes to those with fewer. Little of the literature included consideration of resource and training implications of interventions, apart from a small number of studies which reported the value of skilled clinicians. There seems to be a research gap around aspects of process evaluation such as intervention fidelity; practitioner specific effects, acceptability, and feasibility.

The challenge in establishing what a 'good outcome' following intervention should be is also a key issue for the field. While a sizeable body of studies included in this review reported effectiveness in terms of reduction in the overt frequency or severity of stuttering it is debateable how significant a reduction of, for example, two to three syllables per 100 syllables might be for the everyday functioning of a PWS, or indeed whether this reduction in overt stuttering level was the issue of most concern to a PWS. While there is some evidence of increasing involvement of PWS in the determination of outcomes, the field remains dominated by measures of overt stuttering behaviours. Only a small number of papers (all relating to the LP) considered whether interventions could have a potential adverse impact. Studies describing speech motor interventions often considered the effect on speech naturalness, however rating was often carried out by an independent listener, with few including rating or perceptions from the PWS. Further understanding regarding how and to what degree intervention outcomes relate to the everyday lives of PWS is urgently needed.

Limitations of the study

This work used established systematic review methods to identify and examine published evidence in a transparent and replicable process. We examined a large body of work, however the review may be limited by excluding studies using case studies/series and survey designs. While we recognize that these designs can be of value, they are inherently at high risk of bias with limited external validity and with stronger designs available, the decision was made to exclude them from this review. There may however be useful data within these studies which could have contributed to the review. The body of work that we included encompassed both studies that

we categorized as being at higher risk of bias, and those at lower risk. We considered whether to use quality criterion as a basis for rejection, however this would have precluded analysis of a large quantity of the literature. In consideration of the results however it is important to recognize the study quality limitations. The heterogeneous nature of the literature and variability in outcomes reported precluded any meaningful statistical comparison of intervention effectiveness.

Conclusions

This systematic review of evidence for the effectiveness of interventions for PWS across all intervention types and populations has highlighted evidence of effectiveness across the range of treatments available to PWS, and where future research priorities lie. The field has a good body of small sample baseline follow-up investigations suggesting that alternative study designs are required in the future such as research comparing interventions While the literature currently has a tendency for focusing on demonstrating that a particular intervention is effective, the evidence base suggests a need instead to investigate how and why therapy works, and in particular a need to further investigate individual variation in effectiveness. Factors such as severity of stuttering and length of onset have been suggested as being influential in outcomes however, much of the variance in response has not been explained. While different studies continue to use varied measures of stuttering, comparison between them remains challenging. While measures of overt stuttering behaviours continue to dominate evaluation, the establishment of agreed core outcomes which are comparable between studies and of importance and relevance to PWS seems to be an urgent priority.

Acknowledgements

References

ALLEN, C. R., 2011, The use of email as a component of adult stammering therapy: a preliminary report. *Journal of Telemedicine and Telecare*, **17**, 163–167.

AMSTER, B. J. and KLEIN, E. R., 2008, Perfectionism in people who stutter: preliminary findings using a modified cognitive–behavioral treatment approach. *Behavioural and Cognitive Psychotherapy*, **36**, 35–40.

ANDREWS, C., O'BRIAN, S., HARRISON, E., ONSLOW, M., PACKMAN, A. and MENZIES, R., 2012, Syllable-timed speech treatment for school-age children who stutter: a phase I trial. *Language, Speech and Hearing Services in the Schools*, **43**, 359–369.

ANDREWS, G. and CUTLER, J., 1974, Stuttering therapy: the relationship between changes in symptom level and attitudes. *Journal of Speech and Hearing Disorders*, **38**, 312–319.

ANTIPOVA, E. A. PURDY, S. C. BLAKELEY, M. and WILLIAMS, S., 2008, Effects of altered auditory feedback (AAF) on stuttering frequency during monologue speech production. *Journal of Fluency Disorders*, **33**, 74–90.

ARMSON, J. and KIEFTE, M., 2008, The effect of SpeechEasy on stuttering frequency, speech rate, and speech naturalness. *Journal of Fluency Disorders*, **33**, 120–134.

ARMSON, J., KIEFTE, M., MASON, J. and DE CROOS, D., 2006, The effect of SpeechEasy on stuttering frequency in laboratory conditions. *Journal of Fluency Disorders*, **31**, 37–52.

ARMSON, J. and STUART, A., 1998, Effect of extended exposure to frequency-altered feedback on stuttering during reading and monologue. *Journal of Speech, Language and Hearing Research*, **41**, 479–490.

BAUMEISTER, H., CASPAR, F. and HERZIGER, F., 2003, Treatment outcome study of the stuttering therapy summer camp 2000 for children and adolescents. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, **53**, 455–463.

BEILBY, J. M., BYRNES, M. L. and YARUSS, J. S., 2012, Acceptance and Commitment Therapy for adults who stutter: psychosocial adjustment and speech fluency. *Journal of Fluency Disorders*, **37**, 289–299.

BERKOWITZ, M., COOK, H. and HAUGHEY, M. J., 1994, A non-traditional fluency program developed for the public school setting. *Language, Speech, and Hearing Services in the Schools*, **25**, 94–99.

BLOCK, S., INGHAM, R. J. and BENCH, R. J., 1996, The effects of the Edinburgh Masker on stuttering. *Australian Journal of Human Communication Disorders*, **24**, 11–18.

BLOCK, S., ONSLOW, M., PACKMAN, A. and 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.

BLOCK, S., ONSLOW, M., PACKMAN, A., GRAY, B. and DACAKIS, G., 2005, Treatment of chronic stuttering: outcomes from a student training clinic. *International Journal of Language and Communication Disorders*, **40**, 55–66.

BLOCK, S., ONSLOW, M., ROBERTS, R. and WHITE, S., 2004, Control of stuttering with EMG feedback. *Advances in Speech Language Pathology*, **6**, 100–106.

BLOMGREN, M., 2010, Stuttering treatment for adults: an update on contemporary approaches. *Seminars in Speech and Language*, **31**, 272–282.

BLOMGREN, M., 2013, Behavioral treatments for children and adults who stutter: a review. *Psychology Research and Behavior Management*, **6**, 9–19.

BLOMGREN, M., ROY, N., CALLISTER, T. and MERRILL, R. M., 2005, Intensive stuttering modification therapy: a multidimensional assessment of treatment outcomes. *Journal of Speech, Language and Hearing Research*, **48**, 509–523.

BLOOD, G. W., 1995, A behavioral–cognitive therapy program for adults who stutter: computers and counseling. *Journal of Fluency Disorders*, **28**, 165–180.

BLOODSTEIN, O., 1987, *A Handbook on Stuttering*, 4th edn (Chicago, IL: National Easter Seal Society).

BOAZ, A. and ASHBY, D., 2003, *Fit for Purpose? Assessing Research Quality for Evidence Based Policy and Practice* (London: ESRC UK Centre for Evidence Based Policy and Practice).

BOBERG, E. and KULLY, D., 1994, Long-term results of an intensive treatment program for adults and adolescents who stutter. *Journal of Speech, Language and Hearing Research*, **37**, 1050–1059.

BONELLI, P., DIXON, M., RATNER, N. B. and ONSLOW, M., 2000, Child and parent speech and language following the Lidcombe Programme of early stuttering intervention. *Clinical Linguistics and Phonetics*, **14**, 427–446.

BOTHE, A., DAVIDOW, J. H., BRAMLETT, R. and INGHAM, R., 2006a, Stuttering treatment research 1970–2005: I. A systematic review incorporating trial quality assessment of behavioural, cognitive, and related approaches. *American Journal of Speech–Language Pathology*, **15**, 321–341.

BOTHE, A., DAVIDOW, J. H., BRAMLETT, R. and INGHAM, R., 2006b, Stuttering treatment research 1970–2005: II. A systematic review incorporating trial quality assessment of pharmacological approaches. *American Journal of Speech–Language Pathology*, **15**, 342–352.

BRAY, M. and JAMES, S., 2009, An evaluation of a telephone assistive device (TAD) for people who stutter. *International Journal of Speech–Language Pathology*, **11**, 54–60.

BRAY, M. A. and KEHLE, T. J., 1998, Self-modeling as an intervention for stuttering. *School Psychology Review*, **27**, 587–598.

CAREY, B., O'BRIAN, S., ONSLOW, M., BLOCK, S., JONES, M. and PACKMAN, A., 2010, Randomized controlled non-inferiority trial of a telehealth treatment for chronic stuttering: the Camperdown Program. *International Journal of Language and Communication Disorders*, **45**, 108–120.

CENTRE FOR REVIEWS AND DISSEMINATION (CRD), 2008, *Systematic Review: CRD's Guidance for Undertaking Systematic Reviews in Healthcare* (New York, NY: CRD).

COCHRANE COLLABORATION, 2011, *Cochrane Handbook for Systematic Reviews of Interventions, Cochrane Collaboration, Version 5.1.0 March 2011* (available at: http://www.cochrane-handbook.org\051 (accessed on 4 September 2012).

COCOMAZZO, N., BLOCK, S., CAREY, B., O'BRIAN, S., ONSLOW, M., PACKMAN, A. and LVERARCH, L., 2012, Camperdown Program for adults who stutter: a student training clinic Phase I trial. *International Journal of Language and Communication Disorders*, **47**, 365–372.

CRAIG, A., HANCOCK, K., CHANG, E., MCCREADY, C., SHEPLEY, A., MCCAUL, A., COSTELLO, D., HARDING, S., KEHREN, R., MASEL, C. and REILLY, K., 1996, A controlled clinical trial for stuttering in persons aged 9 to 14 years. *Journal of Speech and Hearing Research*, **39**, 808–826.

CRAIG, A., HANCOCK, K. and COBBIN, D., 2002, Managing adolescents who relapse following treatment for stuttering. *Asia Pacific Journal of Speech, Language and Hearing*, **7**, 79–91.

CRAIG, A. R., FRANKLIN, J. A. and ANDREWS, G., 1984, A scale to measure locus of control of behavior. *British Journal of Medical Psychology*, **57**, 173–180.

CREAM, A., O'BRIAN, S., JONES, M., BLOCK, S., HARRISON, E., LINCOLN, M., HEWAT, S., PACKMAN, A., MENZIES, R. and ONSLOW M., 2010, Randomized controlled trial of video self-modeling following speech restructuring treatment for stuttering. *Journal of Speech, Language and Hearing Research*, **53**, 887–897.

CREAM, A., O'BRIAN, S., ONSLOW, M., PACKMAN, A. and MENZIES, R., 2009, Self-modelling as a relapse intervention following speech-restructuring treatment for stuttering. *International Journal of Language and Communication Disorders*, **44**, 587–599.

DE VEER, S., BROUWERS, A., EVERS, W. and TOMIC, W., 2009, A pilot study of the psychological impact of the mindfulness-based stress reduction program on persons who stutter. *European Psychotherapy*, **9**, 39–56.

DRUCE, T., DEBNEY, S. and BYRT, T., 1997, Evaluation of an intensive treatment program for stuttering in young children. *Journal of Fluency Disorders*, **22**, 169–186.

ELLIOTT, A. J., MILTENBERGER, R. G., RAPP, J., LONG, E., S. and MCDONALD, R., 1998, Brief application of simplified habit reversal to treat stuttering in children. *Journal of Behavior Therapy and Experimental Psychiatry*, **29**, 289–302.

FEMRELL, L., AVALL, M. and LINDSTROM, E., 2012, Two-year follow-up of the Lidcombe Program in ten Swedish-speaking children. *Folia Phoniatrica et Logopedica*, **64**, 248–253.

FOUNDAS, A. L., MOCK, J. R., COREY, D. M., GOLOB, E., J. and CONTURE, E. G., 2013, The SpeechEasy device in stuttering and nonstuttering adults: fluency effects while speaking and reading. *Brain and Language*, **126**, 141–150.

FRANKEN, M. C., BOVES, L., PETERS, H. F. and WEBSTER, R. L., 1992, Perceptual evaluation of the speech before and after fluency shaping stuttering therapy. *Journal of Fluency Disorders*, **17**, 223–241.

FRANKEN, M. C., KIELSTRA-VAN DER SCHALK, C. J. and BOELENS, H., 2005, Experimental treatment of early stuttering: a preliminary study. *Journal of Fluency Disorders*, **30**, 189–199.

FRANKLIN, D. E., TAYLOR, C. L., HENNESSEY, N. W. and BEILBY, J. M., 2008, Investigating factors related to the effects of time-out on stuttering in adults. *International Journal of Language and Communication Disorders*, **43**, 283–299.

GAGNON, M. and LADOUCEUR, R., 1992, Behavioral treatment of child stutterers: replication and extension. *Behavior Therapy*, **23**, 113–129.

GALLOP, R. F. and RUNYAN, C. M., 2012, Long-term effectiveness of the SpeechEasy fluency-enhancement device. *Journal of Fluency Disorders*, **37**, 334–343.

GRANT, M. J. and BOOTH, A., 2009, A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, **26**, 91–108.

HANCOCK, K. and CRAIG, A., 1998, Predictors of stuttering relapse one year following treatment for children aged 9 to 14 years. *Journal of Fluency Disorders*, **23**, 31–48.

HANCOCK, K. and CRAIG, A., 2002, The effectiveness of re-treatment for adolescents who stutter. *Asia Pacific Journal of Speech, Language and Hearing*, **7**, 138–156.

HANCOCK, K., CRAIG, A., MCCREADY, C., MCCAUL, A., COSTELLO, D., CAMPBELL, K. and GILMORE, G., 1998, Two- to six-year controlled-trial stuttering outcomes for children and adolescents. *Journal of Speech, Language and Hearing Research*, **41**, 1242–1252.

HARRIS, V., ONSLOW, M., PACKMAN, A., HARRISON, E. and MENZIES, R., 2002, An experimental investigation of the impact of 24 the Lidcombe Program on early stuttering. *Journal of Fluency Disorders*, **27**, 203–213.

HARRISON, E., ONSLOW, M. and MENZIES, R., 2004, Dismantling the Lidcombe Program of early stuttering intervention: verbal contingencies for stuttering and clinical measurement. *International Journal of Language and Communication Disorders*, **39**, 257–267.

HASBROUCK, J. M., 1992, FAMC Intensive Stuttering Treatment Program: ten years of implementation. *Military Medicine*, **157**, 244–247.

HERDER, C. HOWARD, C., NYE, C. and VANRYCKEGHEM, M., 2006, Effectiveness of behavioural stuttering treatment: a systematic review and meta-analysis. *Contemporary Issues in Communication Science and Disorders*, **33**, 61–73.

HEWAT, S., ONSLOW, M., PACKMAN, A. and O'BRIAN, S., 2006, A phase II clinical trial of self-imposed time-out treatment for stuttering in adults and adolescents. *Disability Rehabilitation*, **15**, 33–42.

HUDOCK, D. and KALINOWSKI, J., 2014, Stuttering inhibition via altered auditory feedback during scripted telephone conversations. *International Journal of Language and Communication Disorders*, **49**, 139–147.

HUINCK, W. J., LANGEVIN, M., KULLY, D., GRAAMANS, K., PETERS, H. F. and HULSTIJN, W., 2006, The relationship between pre-treatment clinical profile and treatment outcome in an integrated stuttering program. *Journal of Fluency Disorders*, **31**, 43–63.

INGHAM, R. J., KILGO, M., INGHAM, J. C., MOGLIA, R., BELKNAP, H. and 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.

INGHAM, R. J., WANG, Y., INGHAM, J. C., BOTHE, A. K. and GRAFTON, S. T., 2013, Regional brain activity change predicts responsiveness to treatment for stuttering in adults. *Brain and Language*, **127**, 510–519.

IRANI, F, GABEL, R., DANIELS, D. and HUGHES, S. T., 2012, The long-term effectiveness of intensive stuttering therapy: a mixed methods study. *Journal of Fluency Disorders*, **37**, 164–178.

IVERACH, L., JONES, M., O'BRIAN, S., BLOCK, S. LINCOLN, M., HARRISON, E., HEWAT S., CREAM A., MENZIES R.G., PACKMAN A. and ONSLOW M., 2009, The relationship between mental health disorders and treatment outcomes among adults who stutter. *Journal of Fluency Disorders*, **34**, 29–43.

JOHNSON, M., BAXTER, S., BLANK, L., CANTRELL, A., BRUMFITT, S., ENDERBY, P. and GOYDER, E. (2015) The state of the art in non-pharmacological interventions for developmental stuttering. Part 2: qualitative evidence synthesis of views and experiences. *International Journal of Language and Communication Disorders*.

JONES, M., ONSLOW, M., PACKMAN, A., O'BRIAN, S., HEARNE, A., WILLIAMS, S., ORMOND, T. and SCHWARZ, I., 2008, Extended follow-up of a randomized controlled trial of the Lidcombe Program of Early Stuttering Intervention. *International Journal of Language and Communication Disorders*, **43**, 649–661.

JONES, M., ONSLOW, M., HARRISON, E. and 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.

JONES, M., ONSLOW, M., PACKMAN, A., WILLIAMS, S., ORMOND, T., SCHWARZ, I. and GEBSKI V., 2005, Randomised controlled trial of the Lidcombe programme of early stuttering intervention. *British Medical Journal*, **24**, 659.

KAYA, Y., 2011, The effect of regulator and stiffener exercises for speaking mechanics on the stuttering education and therapy. *Energy Education Science and Technology Part B—Social and Educational Studies*, **3**, 253–268.

KAYA, Y. and ALLADIN, A., 2012, Hypnotically assisted diaphragmatic exercises in the treatment of stuttering: a preliminary investigation. *International Journal of Clinical and Experimental Hypnosis*, **60**, 175–205.

KINGSTON, M., HUBER, A., ONSLOW, M., JONES, M. and PACKMAN, A., 2003, Predicting treatment time with the Lidcombe Program: replication and meta-analysis. *International*

Journal of Language and Communication Disorders, **38**, 165–177.

KOUSHIK, S., HEWAT, S., SHENKER, R. C., JONES, M. and ONSLOW, M., 2011, North American Lidcombe Program file audit: replication and meta-analysis. *International Journal of Speech–Language Pathology*, **13**, 301–307.

KOUSHIK, S., SHENKER, R. and ONSLOW, M., 2009, Follow-up of 6–10-year-old stuttering children after Lidcombe program treatment: a phase I trial. *Journal of Fluency Disorders*, **34**, 279–290.

LAIHO, A. and KLIPPI, A., 2007, Long- and short-term results of children's and adolescents' therapy courses for stuttering. *International Journal of Language and Communication Disorders*, **42**, 367–382.

LANGEVIN, M. and BOBERG, E., 1993, Results of an intensive stuttering therapy program. *Journal of Speech–Language Pathology and Audiology*, **17**, 158–166.

LANGEVIN, M. and BOBERG, E., 1996, Results of intensive stuttering therapy with adults who clutter and stutter. *Journal of Fluency Disorders*, **21**, 315–327.

LANGEVIN, M., HUINCK, W. J., KULLY, D., PETERS, H. F., LOMHEIM, H. and 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.

LANGEVIN, M., KULLY, D., TESHIMA, S., HAGLER, P. and NARASIMHA PRASAD, N. G., 2010, Five-year longitudinal treatment outcomes of the ISTAR Comprehensive Stuttering Program. *Journal of Fluency Disorders*, **35**, 123–140.

LATTERMANN, C., EULER, H. A. and NEUMANN, K., 2008, A 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.

LAWSON, R., PRING, T. and FAWCUS, M., 1993, The effects of short courses in modifying the attitudes of adult and adolescent stutterers to communication. *European Journal of Disorders of Communication*, **28**, 299–308.

LEAHY, M. M. and COLLINS, G., 1991, Therapy for stuttering: experimenting with experimenting. *Irish Journal of Psychological Medicine*, **8**:37–39.

LEWIS, C., PACKMAN, A., ONSLOW, M., SIMPSON, J. M. and JONES, M., 2008, A 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. A. and ONSLOW, M., 1997, Long-term outcome of early intervention for stuttering. *American Journal of Speech–Language Pathology*, **6**, 51–58.

LINCOLN, M., ONSLOW, M., LEWIS, C. and WILSON, L., 1996, A clinical trial of an operant treatment for school-age children who stutter. *American Journal of Speech–Language Pathology*, **5**, 73–85.

LINCOLN, M., PACKMAN, A. and ONSLOW M., 2006, Altered auditory feedback and the treatment of stuttering: a review. *Journal of Fluency Disorders*, **31**, 71–89.

LUTZ, C., 2009, The Hamburg Workshop for Parents of Stuttering Children (HAWESK). *Forum Logopadie*, **23**, 6–14.

MALLARD, A. R., 1998, Using problem-solving procedures in family management of stuttering. *Journal of Fluency Disorders*, **23**, 127–135.

MENZIES, R. G., O'BRIAN, S., ONSLOW, M., PACKMAN, A., ST CLARE, T. and BLOCK, S., 2008, An experimental clinical trial of a cognitive–behavior therapy package for chronic stuttering. *Journal of Speech, Language and Hearing Research*, **51**, 1451–1464.

MILLARD, S. K., EDWARDS, S. and COOK, F. M., 2009, Parent–child interaction therapy: adding to the evidence. *International Journal of Speech–Language Pathology*, **11**, 61–76.

MILLARD, S. K., NICHOLAS, A. and COOK, F. M., 2008, Is parent–child interaction therapy effective in reducing stuttering? *Journal of Speech, Language and Hearing Research*, **51**, 636–650.

MILLER, B. and GUITAR, B., 2009, Long-term outcome of the Lidcombe Program for early stuttering intervention. *American Journal of Speech–Language Pathology*, **18**, 42–49.

NILSEN, C. H. and RAMBERG, C., 1999, Evaluation of a Scandinavian Intensive Program for Stuttering in Adolescence. *Logopedia Phoniatrica Vocology*, **24**, 66–75.

NYE, C., VANRYCKEGHEM, M., SCHWARTZ, J. B., HERDER, C., TURNER III, H. M. and HOWARD, C., 2013, Behavioral stuttering interventions for children and adolescents: a systematic review and meta-analysis. *Journal of Speech, Language and Hearing Research*, **56**, 921–932.

O'BRIAN, S., IVERACH, L., JONES, M., ONSLOW, M., PACKMAN, A. and MENZIES, R., 2013, Effectiveness of the Lidcombe Program for early stuttering in Australian community clinics. *International Journal of Speech–Language Pathology*, **15**, 593–603.

O'BRIAN, S., ONSLOW, M., CREAM, A. and PACKMAN, A., 2003, The Camperdown Program: outcomes of a new prolonged-speech treatment model. *Journal of Speech, Language and Hearing Research*, **46**, 933–946.

O'BRIAN, S., PACKMAN, A. and ONSLOW, M., 2008, Telehealth delivery of the Camperdown Program for adults who stutter: a phase I trial. *Journal of Speech, Language and Hearing Research*, **51**, 184–195.

O'DONNELL, J. J., ARMSON, J. and KIEFTE, M., 2008, The effectiveness of SpeechEasy during situations of daily living. *Journal of Fluency Disorders*, **33**, 99–119.

ONSLOW, M., ANDREWS, C. and LINCOLN, M., 1994, A control/experimental trial of an operant treatment for early stuttering. *Journal of Speech, Language and Hearing Research*, **37**, 1244–1259.

ONSLOW, M., COSTA, L., ANDREWS, C., HARRISON, E. and PACKMAN, A., 1996, Speech outcomes of a prolonged-speech treatment for stuttering. *Journal of Speech, Language and Hearing Research*, **39**, 734–749.

ONSLOW, M., COSTA, L. and RUE, S., 1990, Direct early intervention with stuttering: some preliminary data. *Journal of Speech and Hearing Disorders*, **55**, 405–416.

ONSLOW, M., HAYES, B., HUTCHINS, L. and NEWMAN, D., 1992, Speech naturalness and prolonged-speech treatments for stuttering: further variables and data. *Journal of Speech, Language and Hearing Research*, **35**, 274–282.

ONSLOW, M., STOCKER, S., PACKMAN, A. and MCLEOD, S., 2002, Speech timing in children after the Lidcombe Program of early stuttering intervention. *Clinical Linguistics and Phonetics*, **16**, 21–33.

PAPE-NEUMANN, J., 2004, Results of the test-phase of PEVOS, program for the evaluation of stuttering therapy. *Forum Logopädie*, **18**, 18–22.

POLLARD, R., ELLIS, J. B., FINAN, D. and RAMIG, P. R., 2009, Effects of the SpeechEasy on objective and perceived aspects of stuttering: a 6-month, phase I clinical trial in naturalistic environments. *Journal of Speech, Language and Hearing Research*, **52**, 516–533.

RATYNSKA, J., SZKIELKOWSKA, A., MARKOWSKA, R., KURKOWSKI, M., MULARZUK, M. and SKARZYNSKI, H., 2012, Immediate speech fluency improvement after application of the Digital Speech Aid in stuttering patients. *Medical Science Monitor*, **18**, 9–12.

REDDY, R. P., SHARMA, M. P. and SHIVASHANKAR, N., 2010, Cognitive behavior therapy for stuttering: a case series. *Indian Journal of Psychological Medicine*, **32**, 49–53.

RILEY, G. D., 1994, *Stuttering Severity Instrument for Children and Adults—Third Edition (SSI-3)* (Austin, TX: Pro-Ed).

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

ROSENBERGER, S., SCHULET, K. and METTEN, C., 2007, Intensive stuttering therapy—first results of an evaluation study. *Forum Logopädie*, **21**, 20–25.

ROUSSEAU, I., PACKMAN, A., ONSLOW, M., HARRISON, E. and JONES, M., 2007, An investigation of language and phonological development and the responsiveness of preschool-age children to the Lidcombe Program. *Journal of Communication Disorders*, **40**, 382–397.

RYAN, B. P. and VAN KIRK, R. 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.

SCHARDT, M., ADAMS M. B. OWENS, T., KEITZ, S. and FONTELO, P., 2007, Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making*, **7**, 16. doi:10.1186/1472-6947-7-16

SICOTTE, C., LEHOUX, P., FORTIER-BLANC, J. and LEBLANC Y., 2003, Feasibility and outcome evaluation of a telemedicine application in speech–language pathology. *Journal of Telemedicine and Telecare*, **9**, 253–258.

SIDAVI, A. and FABUS, A., 2010, A review of stuttering intervention approaches for preschool-age and elementary school-age children. *Contemporary Issues in Communication Science Disorders*, **37**, 14–26.

SMITS-BANDSTRA, S. M. and YOVETICH, W. S., 2003, Treatment effectiveness for school-age children who stutter. *Journal of Speech–Language Pathology and Audiology*, **27**, 125–133.

SPENCER, L., RITCHIE, J., LEWIS, J. and DILLON, L., 2003, *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence* (London: National Centre for Social Research).

STEWART, T., 1996, A further application of the Fishbein and Ajzen model to therapy for adult stammerers. *European Journal of Disorders of Communication*, **31**, 445–464.

STIDHAM, K. R., OLSON, L., HILLBRATT, M. and SINOPOLI, T., 2006, A new antistuttering device: treatment of stuttering using bone conduction stimulation with delayed temporal feedback. *Laryngoscope*, **116**, 1951–1955.

STUART, A., KALINOWSKI, J., RASTATTER, M., SALTUKLAROGLU, T. and DAYALU, V., 2004, Investigations of the impact of altered auditory feedback in-the-ear devices on the speech of people who stutter: initial fitting and 4-month follow-up. *International Journal of Language and Communication Disorders*, **39**, 93–113.

STUART, A., KALINOWSKI, J., SALTUKLAROGLU, T. and GUNTUPALLI, V. K., 2006, Investigations of the impact of altered auditory feedback in-the-ear devices on the speech of people who stutter: one-year follow-up. *Disability Rehabilitation*, **30**, 757–765.

TRAJKOVSKI, N., ANDREWS, C., ONSLOW, M., O'BRIAN, S., PACKMAN, A. and MENZIE, S. R., 2011, A phase II trial of the Westmead Program: syllable-timed speech treatment for preschool children who stutter. *International Journal of Speech–Language Pathology*, **13**, 500–509.

UNGER, J. P., GLUCK, C. W. and CHOLEWA, J., 2012, Immediate effects of AAF devices on the characteristics of stuttering: a clinical analysis. *Journal of Fluency Disorders*, **37**, 122–134.

VAN BORSEL, J., REUNES, G. and VAN DEN BERGH, N., 2003, Delayed auditory feedback in the treatment of stuttering: clients as consumers. *International Journal of Language and Communication Disorders*, **38**, 119–129.

VON GUDENBERG, A. W., 2006, Kassel stuttering therapy: evaluation of a computer aided therapy. *Forum Logopadie*, **20**, 6–11.

VON GUDENBERG, A. W., NEUMANN, K. and EULER, H. A., 2006, The Kassel Stuttering Therapy fills the gap of therapy options for older children. *Forum Logopadie*, **20**, 24–29.

WAGAMAN, J. R., MILTENBERGER, R. G. and ARNDORFER, R. E., 1993, Analysis of a simplified treatment for stuttering in children. *Journal of Applied Behavior Analysis*, **26**, 53–61.

WAGAMAN, J. R., MILTENBERGER, R. G. and WOODS, D., 1995, Long-term follow-up of a behavioral treatment for stuttering in children. *Journal of Applied Behavior Analysis*, **28**, 233–234.

WARD, D., 1992, Outlining semi-intensive fluency therapy. *Journal of Fluency Disorders*, **17**, 243–255.

WILLE, A., 1999, Bioresonance therapy (biophysical information therapy) in stuttering children. *Forschende Komplementarmedizin*, **6(Suppl. 2)**.

WILSON, L., ONSLOW, M. and LINCOLN, M., 2004, Telehealth adaptation of the Lidcombe Program of Early Stuttering Intervention: five case studies. *American Journal of Speech–Language Pathology*, **13**, 81–93.

WOODS, S., SHEARSBY, J., ONSLOW, M. and BURNHAM, D., 2002, Psychological impact of the Lidcombe Program of early stuttering intervention. *International Journal of Language and Communication Disorders*, **37**, 31–40.

WOOLF, G., 1967, The assessment of stuttering as struggle, avoidance, and expectancy. *British Journal of Disorders of Communication*, **2**, 158–171.

YAIRI, E. and AMBROSE, N., 1992, A longitudinal study of stuttering in children: a preliminary report. *Journal of Speech and Hearing Research*, **35**, 755–760.

YARUSS, J. S., COLEMAN, C. and HAMMER, D., 2006, Treating preschool children who stutter: description and preliminary evaluation of a family-focused treatment approach. *Language, Speech and Hearing Services in the Schools*, **37**, 118–136.

ZIMMERMAN, S., KALINOWSKI, J., STUART, A. and RASTATTER, M., 1997, Effect of altered auditory feedback on people who stutter during scripted telephone conversations. *Journal of Speech, Language and Hearing Research*, **40**, 1130–1144.

Appendix
Quality appraisal for individual studies

Risk of bias: yes/no/unclear

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Allen, 2011	Yes	Yes	No	Yes	Yes	Higher	Small sample, unclear research questions and recruitment justification, poor reporting
Amster and Klein, 2008	Yes	Yes	No	Yes	No	Higher	Small sample, no control, volunteered sample
Andrews *et al.*, 2012	Yes	Yes	Yes	No	No	Higher	Small sample, no control, volunteered sample
Antipova *et al.*, 2008	Yes	Yes	No	Yes	No	Higher	Small sample, no control, volunteered sample
Armson and Stuart, 1998	Yes	Yes	No	Yes	No	Higher	Small sample, experimental design with no follow-up (FU). Single session tests, Kappa scoring methods not described or reliability/results in detail

	1. Selection bias: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. Performance bias: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. Attrition bias: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. Detection bias: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. Reporting bias: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Armson et al., 2006	Yes	Yes	No	Yes	No	Higher	Small sample, experimental design with no FU
Armson and Kiefte, 2008	Yes	Yes	No	Yes	No	Higher	Mid-sized sample experimental design with no FU. First 31 people taken into study
Baumeister et al., 2003	Yes	Yes	Yes	No	No	Higher	Large sample, but no control group. Participants showed different severity of disorder which influenced results. Some participants dropped out or were not assessed at baseline
Beilby et al., 2012	No	Unclear	No	No	No	Lower	Unclear if raters were blinded to time point, 3-month FU
Berkowitz et al., 1994	Yes	Yes	No	No	No	Higher	Very small sample, no control, no blinding in assessment, self-reports used
Block et al., 2004	No	No	No	Yes	Yes	Higher	Sample 12, 5-min conversation 5-min reading. Unclear who recorded away from clinic. Basic results for post-treatment periods, 3-month FU, limited analysis

Continued

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Block *et al*, 2005	No	No	No	No	No	Lower	Large sample, self-report inventory used at 35-year FU with 87% of sample response rate. Unclear length of speech sample
Block *et al*, 2006	No	No	No	No	No	Lower	Same study as 2005 paper with further examination of variables
Block *et al*, 1996	No	No	No	Yes	No	Higher	Larger sample, no dropout, immediate measurement during intervention, experimental setting, 5-min samples
Blomgren *et al.*, 2005	Yes	Unclear	No	No	No	Lower	Sample 19. Some use of self-reported outcome measures post-study. Sample 4 min of speech, unclear if rater blinded, 6-month FU
Blood, 1995	Yes	Yes	No	Yes	Yes	Higher	Extremely small sample. Flawed recruitment. Use of self-reported outcomes

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Boberg and Kully, 1994	No	Unclear	No	Yes	Yes	Higher	Sample 42, no control. Telephone call sample 2 min. Unclear if raters blind to time point, percentage change reported
Bonelli *et al.*, 2000	Yes	No	No	No	Yes	Higher	Sample of 9 selected from earlier study, no pooling of data reported by individual only
Bray and James, 2009	Yes	Yes	No	Yes	Yes	Higher	Small sample, use of self-reported outcomes
Bray and Kehle, 1998	Yes	Yes	No	Yes	Yes	Higher	Small sample (4), volunteers. Content of speech sample and listener varied between individuals and time, descriptive data by individual only
Carey *et al.*, 2010	No	No	No	No	No	Lower	20 per trial arm, with 75% loss to FU, 12-month FU, 10-min recording via telephone, blinded assessment

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Cocomazzo et al., 2012	No	No	Yes	Unclear	No	Lower	12 participants and dropouts, blinded rating, beyond clinic recordings made by participant but asked to make only one, 12-month FU
Craig et al., 1996	No	No	No	Yes	No	Lower	Larger sample, raters blinded 12-month FU, 5-min speech samples
Craig et al., 2002	Yes	Yes	No	Yes	No	Higher	Small sample (6) selected from previous study, 2-year FU, descriptive data for individuals only, home measure potential for bias
Cream et al., 2009	Yes	No	Yes	Yes	No	Higher	Sample of 10, 5-min recordings, use of some self-reported outcomes, 2 dropouts in small sample, blinded assessor, immediate post-assessment
Cream et al., 2010	No	No	No	No	No	Lower	Randomized sample with acceptable dropout rate, blinded assessment, 6-month FU

Continued

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
De Veer et al., 2009	Yes	Yes	No	Yes	No	Higher	Selection of potential participants by researchers. No detail of randomization. No measure of fluency, self-report measures only
Druce et al., 1997	Yes	No	No	Yes	No	Lower	Sample 15 with adequate FU, 2-min speech sample, raters blinded
Elliott et al., 1998	Yes	Yes	No	No	Yes	Higher	Small sample (5) 5-min conversation sample, reported by individuals, limited analysis
Femrell et al., 2012	Yes	Yes	Yes	Yes	No	Higher	Sample 10 with 2 loss to FU, 10-min conversation, assessed by clinician, not blinded
Foundas et al., 2013	No	Unclear	No	Unclear	No	Higher	Sample 24 (10 control) with control and experimental conditions. Immediate outcomes, length of sample not reported. Unclear whether blinded

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Franken *et al.*, 1992	No	No	No	Yes	No	Higher	Fair sized sample, (32) 6-month FU, rating scale used, blinded assessment, control normal speakers, 5-min recordings, purpose to compare normal to post-intervention rather than evaluate interventions per se
Franken *et al.*, 2005	No	No	Yes	Yes	No	Lower	Small sample randomized to 1 of 2 arms. Loss to FU, recordings made by parents, blinded assessors
Franklin *et al.*, 2008	No	Yes	No	Unclear	No	Higher	Larger sample, however, participants were not randomized to each condition, assessment carried out by authors not blinded, immediate FU, sample 1500 syllables
Gagnon and Ladouceur, 1992	Yes	Unclear	No	Unclear	Yes	Higher	Small samples used in separate studies. Data presented by individual, lack of clarity regarding data collection and evaluation

Continued

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Gallop and Runyan, 2012	Yes	Unclear	Unclear	Yes	Yes	Higher	Sample 11 participants no control, no explanation of recruitment criteria, 15-min telephone samples, not reported if blinded, informal FU
Hancock and Craig, 1998	Yes	No	No	No	No	Lower	Larger sample (77 participants), 12-month FU. Pseudo-randomization, 5-min speech sample, in clinic at distance via phone
Hancock *et al.*, 1998	Yes	No	No	No	No	Lower	Same study as other paper. This paper reports some different outcomes
Hancock and Craig, 2002	Yes	No	No	Yes	No	Lower	Sample (12) selected from earlier study 5-min only speech sample
Harris *et al.*, 2002	No	No	Yes	No	No	Lower	Small study 29 participants, 6 dropped out (21%)
Harrison *et al.*, 2004	No	No	Yes	No	No	Lower	46 participants, 8 dropped out

	1. Selection bias: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ±10 sample	2. Performance bias: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. Attrition bias: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. Detection bias: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. Reporting bias: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Hasbrouck, 1992	Unclear	Yes	No	No	Unclear	Higher	Larger sample, no control, no blinding. Very sparse details given about recruitment, attrition, analysis
Hewat et al., 2006	No	No	Yes	Yes	No	Lower	30 participants recruited, dropout varied from 27% to 40% at different stages of the study, in clinic recording and participant selected recording, rating blinded
Hudock and Kalinowski, 2014	Unclear	No	No	Yes	Yes	Higher	Small study (9) no detail of recruitment, scripted telephone conversations, immediate measurement
Huinck et al., 2006	Unclear	No	No	No	No	Lower	25 participants. No details given about recruitment methods

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Ingham *et al.*, 2013	Yes	Unclear	Yes	Yes	No	Higher	Volunteer participants, 9 of 21 did not complete 3-min monologue, 3-min conversation. Study purpose to compare normal to PWS however contains before and after data. No detail of whether raters blinded, immediate FU, in-clinic rating
Ingham *et al.*, 2001	Yes	No	No	Yes	No	Higher	Small sample (5 participants). Use of some self-reported measures, participants submitted own recording for beyond clinic measure, not detailed whether raters blinded, data reported by participant

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Irani *et al.*, 2012	No	Unclear	No	Yes	No	Higher	Mixed method study some self reported measures, use of inferential statistics, small sample (7) little detail of speech sample analysis
Iverach *et al.*, 2009	No	No	No	Yes	No	Lower	Larger sample (64 participants). Use of some self-reported measures
Jones *et al.*, 2000	No	No	No	No	No	Lower	Large sample (261 children, 4% dropout rate, all explained)
Jones *et al.*, 2005	No	No	No	No	No	Lower	Larger sample (54 children, 13% dropout rate, all explained)
Jones *et al.*, 2008	No	No	Yes	No	No	Lower	This is a 5-year FU of the earlier study 31% of the original treatment group could not be re-contacted, and 68% of the control group

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Kaya and Alladin, 2012	Yes	No	No	Yes	No	Higher	No comparator group. No detail regarding how stuttering occurrences defined. Immediate assessment at final session
Kaya, 2011	Yes	Yes	No	Yes	No	Higher	Assessment via 2-min speech sample only, rating scale measure used very limited
Kingston *et al.*, 2003	No	Yes	No	Yes	No	Higher	Larger sample (78 children). Assessment was done by the clinician/researcher with no blinding. Purpose of paper to examine associations (predict treatment time) rather than outcomes
Koushik *et al.*, 2009	No	No	No	No	Yes	Lower	Sample 12 children, 1 dropped out

Continued

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Koushik *et al.*, 2011	Unclear	Unclear	No	Unclear	No	Higher	Pooled data from 5 clinical sites. Larger sample (134 participants in final analysis). Retrospective file audit. Purpose of study to examine associations rather than evaluate outcomes
Laiho and Klippi, 2007	No	Yes	No	Yes	No	Higher	Sample 21, no control, assessment via video by author, parent-report data for beyond clinic data. Follow-up data only parent report
Langevin and Boberg, 1993	Yes	No	Yes	No	No	Higher	Small sample, high dropout rate (21 participants, 11 dropped out) data reported by individual
Langevin and Boberg, 1996	Yes	No	Yes	No	No	Lower	25 in 1 group, 16 in other. Two-year FU, some loss to FU 2/3-min samples of speech in clinic and via telephone. Raters probably blinded

Study	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Langevin *et al.*, 2006	No	No	No	No	No	Lower	18 participants, no control. Small loss to FU
Langevin *et al.*, 2010	Yes	No	Yes	No	No	Lower	5-year FU of earlier study
Lattermann *et al.*, 2008	No	No	No	No	No	Lower	Sample 46, blinded rating
Lawson *et al.*, 1993	No	No	No	Yes	No	Higher	Self-report measures only used, 1-month FU, some dropout
Leahy and Collins, 1991	Yes	Yes	No	Yes	Yes	Higher	No comparator group, small sample size *n* = 5 Measures taken by student clinician carrying out intervention. Longer FU only for 2. Reporting by individual only
Lewis *et al.*, 2008	No	No	No	No	No	Lower	Small sample (8 in intervention group, 10 in control group)
Lincoln *et al.*, 1996	No	No	No	Yes	Yes	Higher	Sample of 11, high dropout of potential participants (22 recruited). Some pooled data, some reporting of individuals only, 12-month FU, parent-recorded speech data

Continued

	1. Selection bias: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. Performance bias: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. Attrition bias: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. Detection bias: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. Reporting bias: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Lincoln and Onslow, 1997	No	Yes	Yes	Yes	Yes	Higher	Long-term outcomes of earlier studies. Large initial dropout of potential participants. Parents collected speech sample, parent report questionnaire, descriptive data
Lutz, 2009	Yes	Yes	No	Yes	No	Higher	Findings from a workshop for parents using before and after questionnaires
Mallard, 1998	No	Yes	No	Yes	Yes	Higher	Only measure of success was 'is child in S&L therapy 1 year after intervention?' No control group, limited analysis
Menzies *et al.*, 2008	No	No	No	No	No	Lower	Smaller sample—32 participants, 2 dropped out, 16 in each condition
Millard *et al.*, 2008	No	No	Yes	No	No	Higher	Small sample (9) high dropout rate (30%), blinded rating, in clinic assessment, 12-month FU, data by individual only

Continued

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Millard *et al.*, 2009	No	No	Yes	No	Yes	Higher	Small sample (10) high dropout rate, parent-recorded video data. Control group for initial allocation but removed part way, no pooled data descriptive statistics only
Miller and Guitar, 2009	No	No	No	No	No	Lower	Long-term FU, many participants at minimum level Only 2 dropouts from 15, limited speech sample
Nilsen and Ramberg, 1999	No	No	No	Yes	No	Higher	2 dropouts from 13, use of some scales with limited scoring and analysis, data reported by individual
O'Brian *et al.*, 2003	No	No	No	Yes	Yes	Higher	5 dropout from 30. Reasonable sample, no comparator, participant-selected recordings, limited statistical analysis
O'Brian *et al.*, 2008	No	No	Yes	No	Yes	Higher	16 of 30 completed. Descriptive analysis
O'Brian *et al.*, 2013	No	No	No	No	No	Lower	No control

Continued

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
O'Donnell *et al.*, 2008	Yes	No	No	No	Yes	Higher	Small sample (7), 5 from previous study who had shown most benefit. Data reported by individual participant
Onslow *et al.*, 1994	Unclear	No	Yes	Yes	No	Higher	High withdrawal for control and intervention, no detail of allocation, audio recordings made by parents, descriptive statistics
Onslow *et al.*, 1990	Yes	Yes	No	No	Yes	Higher	No comparator group, sample size $n = 4$. Presentation of findings via figures only, no grouping of data
Onslow *et al.*, 1992	Unclear	No	No	No	No	Higher	Focuses on speech naturalness data only comparing PWS and non-stutterers, no control group, immediate FU, recruitment process unclear
Onslow *et al.*, 1996	Unclear	No	Yes	No	No	Higher	Data reported for only 18 of 32 recruited

Continued

	1. Selection bias: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. Performance bias: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. Attrition bias: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. Detection bias: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. Reporting bias: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Onslow et al., 2002	Yes	No	No	No	Yes	Higher	Small sample (8) 6 of whom in previous studies. For 2 parent-only recordings. Descriptive data presented by individual only. Purpose of paper to evaluate 1 aspect of intervention outcome
Pape-Neumann, 2004	Yes	Yes	Yes	Yes	Yes	Higher	This is a pilot study which presents data from a range of interventions
Pollard et al., 2009	No	Yes	No	Yes	No	Lower	Sample 11, no dropout, samples collected at non-laboratory locations, non-blinded scoring, immediate outcomes
Ratynska et al., 2012	Yes	Yes	No	Unclear	No	Higher	Large sample found other treatment ineffective, no dropout, no blinding of assessment, incomplete description of data collection

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Reddy *et al.*, 2010	Unclear	Unclear	No	Unclear	Unclear	Higher	Small sample (5), limited reporting of findings beyond description of cases
Riley and Ingham, 2000	Unclear	Unclear	No	Unclear	Yes	Higher	Sample 12, pseudo-randomization, no blinding of assessors, unclear beyond clinic data collection, unclear whether 12 or 6 participants being reported, no reporting of control group outcomes
Rosenberger, 2007	Yes	Yes	Yes	No	No	Higher	Two groups were compared which have an uneven number of participants. Some dropout. Limited blinding and speech measures
Rousseau *et al.*, 2007	No	Yes	No	No	No	Lower	Reasonable sample large proportion of parent-recorded samples, no analysis of dropouts
Ryan and Van Kirk, 1995	No	Yes	Yes	No	No	Lower	Reasonable sample (24, 20 completed all elements). Pseudo-randomization, no blinding of speech evaluation

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Sicotte *et al.*, 2003	Yes	Yes	No	Yes	No	Higher	Sample 6, rating scales only. Purpose of study to evaluate intervention fidelity rather than outcomes
Smits-Bandstra and Yovetitch, 2003	Yes	Unclear	No	Yes	No	Higher	Small groups. Each time point presented separately in table form, limited discussion of trends over time. Six participants had received other intervention immediately prior 3-min sample, assume no blinding, limited statistical analysis
Stewarr, 1996	No	Yes	No	Yes	No	Higher	Reasonable sample (15) no blinding of assessment, 3-min conversation, limited analysis of speech behaviour data, no outside clinic measure, reasonable FU

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Stidham *et al.*, 2006	Yes	Yes	No	No	No	Higher	Volunteers recruited, small sample (9), immediate FU, no blinding as assessment, in laboratory evaluation
Stuart *et al.*, 2004	Unclear	Yes	No	No	No	Higher	No detail of recruitment, non-blinded assessment, in laboratory evaluation, limited speech sample
Stuart *et al.*, 2006	Unclear	Yes (for some elements)	No	Yes	No	Higher	No detail of recruitment, no blinding of assessment for first studies, limited speech sample (300 syllables), reasonable FU, in clinic assessment
Trajkovski *et al.*, 2011	No	No	Yes	Yes	No	Lower	8 of 17 completed, data provided for dropouts, limited pooled data, parent-collected recordings

Continued

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Unger, 2012	Yes	Yes	No	Yes	No	Higher	Volunteer participants, reasonable sample, in clinic only, immediate FU, no blinding of assessment 2 × 5-min monologues
Van Borsel, 2003	Yes	Yes	No	No	No	Higher	Volunteer participants, in-clinic data, no blinding of recordings, small sample (9)
Von Gudenberg, 2006	Yes	Yes	Yes	Yes	Yes	Higher	No clear description of any measurement, participants or methods used; more a evaluation of collected data up to now
Von Gudenberg *et al.*, 2006	Yes	Yes	Yes	Yes	Yes	Higher	Detailed description of the therapy of the therapy reasons why this therapy may be effective and a good approach for young adults. However, presented results are outcomes with no clear description of methodology and limited analysis

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Wagaman *et al.*, 1993	Yes	No	No	No	No	Lower	Follow-up data from study
Wagaman *et al.*, 1995	Yes	No	No	Yes	No	Higher	Small sample; no long-term FU
Ward, 1992	Yes	Yes	No	Unclear	No	Higher	No long-term outcomes. Reports pilot study findings only, small sample
Wille, 1999	Yes	Yes	Yes	Yes	Yes	Higher	No clear description of method, data collection, data analyses; no reference to other literature
Wilson *et al.*, 2004	No	Yes	Yes	No	Yes	Higher	Small sample, dropouts and lack of data
Woods *et al.*, 2002	Yes	No	No	No	No	Lower	Small sample (8), 1-month FU. No speech data, study focuses on cognitive and language assessments

Continued

	1. *Selection bias*: method used to generate the allocation sequence, method used to conceal the allocation sequence. Presence of control, characteristics of participants at baseline, ± 10 sample	2. *Performance bias*: measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample	3. *Attrition bias*: incomplete outcome data, high level of withdrawals from the study. High dropout rate (above 15%)	4. *Detection bias*: accuracy of measurement of outcomes, length of follow-up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up	5. *Reporting bias*: selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting	Overall risk of bias. Lower/higher	Detail of concerns
Yairi and Ambrose, 1992	No	Unclear	Yes	No	No	Higher	Reasonable sample (27) 2-year FU, speech sample small (around 500 words), 6 dropout, unclear whether speech assessors blinded, in clinic data
Yaruss *et al.*, 2006	No	Yes	No	Yes	No	Higher	Sample 17, speech rated by clinician, FU up to 2 years with no dropout, at least 200 word sample unclear how collected, limited analysis
Zimmerman *et al.*, 1997	No	Unclear	No	Yes	No	Higher	Small sample (9), no long-term FU, scripted conversations, unclear if rater blinded