What counts as evidence?

Swimming against the tide: Valuing *both* clinically informed experimentally controlled case series and randomised controlled trials in intervention research

Best, W.[1,2], Sze, W. P.[1], Edmundson, A.[1] & Nickels, L.[2, 3]

1 Language and Cognition, Division of Psychology and Language Sciences, University College London

2 Australian Research Council Centre of Excellence in Cognition and Its Disorders, Macquarie University, Sydney, Australia

3 Department of Cognitive Science, Macquarie University, Sydney, Australia

Corresponding Author:

Wendy Best

Language and Cognition, Division of Psychology and Language Sciences

University College London

2 Wakefield Street

London WC1N 2PF

Email: w.best@ucl.ac.uk

Abstract

Research into intervention with people with speech and language needs often takes the form of single case/case series experimental studies (SCEDs) or randomised controlled trials (RCTs). This paper explores the nature of these designs, including their strengths/weaknesses and highlights the value of understanding the intervention outcomes for individual participants. An online survey gathered information on speech and language therapists' views on their use of the different research designs. We conclude that both research designs are used to inform practice. SCEDs, in particular, are used in developing theories of intervention and informing therapy with individuals. Sound experimental intervention studies of both designs are needed.

What this paper adds: This paper provides an overview from the literature of the strengths and issues for two research designs: RCTs and SCEDs in general and in relation to speech and language intervention research. The results of a survey of clinicians are presented which point to the value of both designs in informing management and intervention.

Despite concerted moves towards evidence-based practice in speech and language therapy/pathology (SLT/P; e.g., Klee, Stringer & Howard, 2009) the gap between research and clinical reality remains wide. In the context of the current economic situation, constraints on health and education funding are likely to increase. With funding becoming increasingly hard to obtain, policy-makers, funders, researchers and clinicians need to consider in depth which studies will be most informative for clinical practice. This paper aims to help researchers and clinicians jointly select the most appropriate research methods for future speech and language intervention research. We include supporting data from a survey of practicing clinicians.

The debate on methodology is not new but is at a junction[1]: Language intervention research can a) swim with the current tide towards large randomised controlled trials (RCTs); b) can swim against this tide by retaining use of other research designs such as experimentally controlled single case and case series approaches, or as we emphasise here, c) can value hybrid designs which can be analysed at group and individual level.

The goals of this paper are to consider the strengths and limitations of RCTs and single cases/case series in demonstrating effectiveness and/or helping understand the mechanisms of change and establish causal relationships between intervention and outcome. We include examples of intervention studies with adults with aphasia and children with primary speech and language needs, and touch upon interventions with people with other communication impairments. The arguments have strong links with, and applicability to, other health related intervention research where the needs are complex and heterogeneous (e.g., psychiatry, psychotherapy). We do not, however, provide detail on the specifics of the designs debated, as these are discussed in detail elsewhere

---

[1] Note that randomised controlled trials and single case experimental designs for many years were not necessarily in competition. However, there has been an increasing (entirely appropriate) move towards evidence based practice, and funding of services is often dependent on 'evidence', The widely held view that RCTs provide 'evidence' and single case experimental designs do not (or only provide very weak evidence) prompted this paper.

(e.g., Ebbels, 2017; Nickels, Best & Howard, 2015; Thompson, 2006), as are guidelines for reporting these studies (e.g., CONSORT Extension for *N*-of-1 Trials (CENT; Shamseer et al., 2015); the risk-of-bias in *N*-of-1 trials (RoBiNT) scale (Tate et al., 2015, 2016); and the Template for Intervention Description and Replication (TIDieR; Hoffmann et al., 2014)). We also acknowledge that there are alternative designs, also aiming to establish treatment effectiveness, which are not covered here, including, for example, regression discontinuity design (see, for an example from the SLT/P field, Dyson, Solity, Best & Hulme, 2018).

Randomised Controlled Trials

Sackett (1995) notes that in 1960, the RCT was an 'oddity' (p. 61), but yet by the time he was writing it was accepted that virtually no drug could enter clinical practice without a demonstration of its efficacy in such clinical trials. RCTs and meta-analyses of RCTs are now placed at the top of hierarchies of medical evidence (Cohen, Stavri, & Hersh, 2004). The medical evidence hierarchy is very influential; hence this view is pervasive in health, increasingly in education, in research and with those who decide policy and funding allocation: other sources of evidence are placed lower down the hierarchy. It is common amongst researchers and clinicians to view other types of research as stepping stones on the way to an RCT. Indeed, there is widespread acceptance of the RCT as the 'gold standard' in answering the question of whether or not a treatment works (e.g., McArthur & Castles, 2017) and of the view that this is where other research should be eventually heading (e.g., Ebbels, 2017). For example, in a review which found a lack of studies of treatment for dysarthria following acquired brain injury in children and adolescents, Morgan and Vogel (2009) conclude that "Efforts should first be directed at modest well-controlled studies to identify likely efficacious treatments that may then be trialled in multi-centre collaborations using quasi-randomised or RCT methodology" (p. 197).

This emphasis on the randomised clinical trial as the best method for determining whether a treatment 'works' is now ubiquitous and applied across healthcare domains and beyond, with (limited) funding often prioritising RCTs. For example, guidance about whether an article is suitable for consideration in the high impact British Medical Journal states that the appropriate study design to answer the question 'Does this treatment work?' is a systematic review or a RCT. For a brief history of the Randomised Control Trial and Evidence Based Medicine, with some links to SLT/P see Rosenbek (2016).

While prevalent and extremely influential the supremacy of the RCT in Evidence Based Medicine is not without its critics. Cohen, Stavri and Hersh (2004) usefully gather, categorise and debate the five main criticisms of Evidence Based Medicine. However, it is also the RCT which is criticised, as this provides the evidence upon which treatments are deemed effective or otherwise. For example, as we note below, Cohen et al. comment that the outcomes of RCTs apply to groups rather than individuals and therefore are limited to predicting average outcomes and only for an 'acceptably similar' group. Given that the hierarchy of evidence puts RCTs at the top, and that many researchers and clinicians view this as the ultimate aim in intervention research, we start by briefly exploring the strengths and weaknesses of RCTs generally, and more specifically in SLT/P research.

Key features of Randomised Controlled Trials

RCTs use a between groups design with random allocation and blinding. Participants are allocated randomly to treatment and control groups on the basis that this should reduce (conscious or unconscious) 'selection bias' on the part of the experimenters. In other words, there will be a reduced likelihood that the groups will differ in characteristics that may influence the results (especially when it is not known which are important): randomisation should result in these characteristics being equally distributed across the groups. However, when the number of

participants in the study is small, and the disorder heterogenous, then this may not be the case: randomisation may give rise to unbalanced groups (e.g., Shamseer et al., 2015).

RCT designs also aim to blind experimenters, participants, and assessors to group membership. For example, in a drug trial, researchers and participants should be unaware (blind) as to whether the tablet being taken is a placebo or a drug; alternatively, in a communication intervention, those assessing the participants should be unaware of (blind to) the group to which participants have been randomly assigned. Blinding aims to reduce systematic differences between groups in the attention that is provided, in exposure to factors other than the interventions of interest and in the way the groups are viewed or evaluated.

Note that it is only those 'well-designed' RCTs, which include the features described above, that should be considered as 'gold standard'. Not all RCTs include these features, and hence may not achieve the goals of, for example, avoiding the possibility of experimenter (conscious or unconscious) bias and/or unbalanced group membership. Unfortunately, however, all too often, RCTs are examined somewhat uncritically, but, as in every research domain, the strength and scope of the claims that can be made is limited by the strength and nature of the design, and its implementation (Dechartres et al., 2017).

Strengths of RCTs

1) Reduction in potential sources of bias

Random assignment and blinding are key elements of RCTs. The random assignment of participants to different groups (intervention and control or comparator intervention) should ensure the experimenters have not introduced any differences between the two groups. It also ensures that other observed (e.g., language scores, age) or unobserved (e.g., motivation, home environment) participant characteristics should be comparable across the groups (Hahs-Vaughn & Nye, 2008).

Furthermore, there are methods available for smaller group studies which ensure groups can be matched on key characteristics (e.g., minimisation, mentioned in Altman et al., 2001, used, for example, by Bishop, Adams & Rosen, 2006). Blinding of assessors and participants also aims to reduce potential bias in the results. Participants and assessors may be blinded to the group they are in (i.e., whether or not they are receiving the intervention of interest) and/or the phase of the study (i.e., intervention or baseline).

2) Generalisation

If a treatment is shown to be effective in an RCT the same findings should generalise to another group of people who meet the same inclusion criteria and who are given the same treatment. An example in the SLT/P field is the work of Onslow and colleagues into the Lidcombe programme for children who stutter (Jones et al., 2005). The findings from this study, where participants were randomly assigned to treatment and control groups, should be generalisable to other groups of children who stutter who meet criteria for entry into the study.

SLT/Ps working with people with communication disorders may, however, question the generalisability of the results of RCTs because of the heterogeneity of clinical populations. Nevertheless, this is very widely accepted as a perceived strength, particularly in medical research. It is also important that random assignment to group should not be confused with random selection of study participants from the population as a whole (Hahs-Vaughn & Nye, 2008) - it is only with random selection that generalisation to a larger population may be possible.

3) Meta-analyses

RCTs can provide an 'effect size' (Cohen, 1988), and the results can be combined across studies regardless of the statistical significance for the individual studies. This can provide a more accurate measure of the 'real' effect of the therapy which can then usefully inform policy and clinical decision

making. Pring (2004) discusses the advantages and limitations of this approach, and cautions that for meta-analyses to be informative, interventions need to be described in sufficient detail to be sure that the analysis is combining studies which truly are providing the same intervention. He notes that this may be possible in some areas of SLT/P intervention such as approaches to dysfluency where "... distinct approaches to therapy exist and studies may more easily be classified and compared" (p. 295). The reporting of RCTs also needs to be sufficiently clear for meta-analyses to be able to evaluate the strength of the design - this is the aim of the CONSORT statements (e.g., Boutron, Moher, Altman, Schulz & Ravaud, 2008; Moher et al., 2010). While meta-analyses are hampered by variability in outcome measures across studies, in some fields attempts are being made to gain agreement for core outcome measures (e.g., for aphasia, Wallace, Worrall, Rose & Le Dorze, 2016).

Limitations of RCTs

1) Findings from RCTs apply to groups on average and not to individuals

It is important to understand that the results of group studies apply to the group. While they may be generalisable to the population who meet the inclusion criteria, the result for the group may NOT apply to all individuals in the group (Shamseer et al., 2015). Even when a trial has shown the intervention is effective some individuals may not benefit at all (e.g., Smith-Lock, Leitão, Prior & Nickels, 2015), or may indeed be harmed by the intervention. Similarly, Cohen et al. (2004) highlight that the findings from clinical trials apply to groups rather than individuals and hence can only predict *average* outcomes for an 'acceptably similar' group. They note that "individuals will respond, to some extent, in their own unique way to a therapy" (p. 40). What is needed, therefore, is to find ways to combine these general findings from group RCTs that may not apply to all individuals in them, with clinical judgement about an individual.

2) Heterogeneity

In some clinical fields clients may share characteristics sufficiently for them to be, in Cohen et al's (2004) terms, 'acceptably similar' in the relevant dimensions. However, in the case of communication difficulties, people's profiles of ability and difficulty differ greatly. For example, within traditional classifications such as 'global aphasia' there is enormous variability: some individuals may use gestures to communicate, others retain a set of phrases and still others may be able to draw to aid communication. Similarly, the label 'developmental language disorder' entails a wide range of profiles, including children with varying receptive and expressive skills, with and without accompanying speech difficulties, with and without pragmatic language difficulties etc. (Bishop, Snowling, Thompson, Greenhalgh, & the CATALISE-2 consortium, 2017). In order to overcome this issue, studies may set stringent inclusion criteria to maximise homogeneity. However, use of highly restrictive criteria in studies of cognitive rehabilitation runs the risk of excluding participants such that the sample is not representative of the population. Moreover, Cicerone (2005) notes that "Since patients who consent to participate in clinical trials are by definition a self-selected group, they may not be representative of patients who either fail to meet enrolment criteria or refuse to participate" (p. 45). They are likely, for example, to have fewer co-morbid health and/or cognitive problems, be more motivated and have better cognitive processing than participants who do not volunteer to participate.

In sum, the heterogeneity of participants with communication disorders, the use of restrictive criteria and the problem of volunteers for research potentially being unrepresentative of the larger population, all limit the likelihood of the results of RCTs being truly generalisable.

3) Control conditions

In attempting to answer questions about the effectiveness of an intervention in an RCT, an appropriate control condition, applied to the control group, is crucial. A great deal of thought needs to be put into the selection of this control condition and the ways in which it differs from the trialled intervention, as the claims that can be made on the basis of an RCT rest on this choice. In studies of new drugs, this control can be an inactive 'placebo' tablet, with participants unaware whether the tablet is the new drug or not. However, for behavioural interventions such as those used in SLT/P, selection of the control condition is a more complex issue and not easily solved.

One approach is to compare outcomes with those of an *untreated control group*, who are only assessed twice at an interval that equals that of the intervention but receive nothing between the assessments. Often this is implemented as a waitlist control group, who receive the treatment later, as this avoids the need to deny participants of treatment. However, an untreated or waitlist control runs the risk of a potentially ineffective treatment being found to be effective. This is because, while there may be improvement of the intervention group compared to the untreated control, this need not be due to the effect of the intervention. Instead it may be due, for example, to placebo effects or Hawthorne effects (the improvement found as a result of the attention of being in a trial, Thompson, 2006).

Hence an alternative approach may be to include a control which also provides many of the 'nonspecific' aspects of the intervention, such as having positive reinforcement from a clinician, or having to concentrate on a task. For example, if evaluating a language focused computer-presented task, the comparator could be a non-linguistic computer task (e.g, Varley et al., 2016). In some aphasia therapy RCTs (e.g., ACT NoW, Bowen et al., 2012) where an intervention is delivered by a clinician, the control condition has been visits by a trained volunteer, to control for positive effects of engagement and communicative interaction with the clinician who delivers the tested intervention. The argument in this case is that a positive outcome indicates therapy 'works' over and

above the effects of contact and conversational interaction as these are equated across groups. Consequently, this allows the conclusion that the 'active' ingredient of therapy is the intervention that was delivered by the clinician. Importantly, however, no difference between the groups cannot be taken as evidence that the speech-language therapy intervention is not effective. By virtue of the volunteers being trained by clinicians in interacting with and facilitating the communication of people with aphasia, they may be delivering many of the elements that are included in aphasia therapy as delivered by clinicians. Once again, this highlights the importance of careful consideration of the conclusions that can be drawn depending on the nature of the control conditions.

There are also studies where two different interventions are compared. One version, commonly used in medicine, is where one group receive 'usual care'. A potential problem here is the quality and frequency of 'usual care'. For example in an RCT of SLT/P intervention with preschool children with language delay, Glogowska, Roulstone, Enderby and Peters (2000) compared progress over a 12-month period for a 'usual care' group and a group who received one 6-month follow up ('watchful waiting'). However, the 'usual care', group children received relatively little intervention: on average eight sessions (range 0-17) over eight months. Using very limited 'standard intervention' is far from ideal. If outcome for the two groups does not differ, what can be concluded? The perceived lack of effectiveness of 'usual care' therapy may simply be due to not enough therapy being provided. If limited amounts of 'usual care' are contrasted with a larger amount of a different target intervention, while the change may result from specific aspects of the intervention, the difference could also simply stem from the difference in the amount of intervention, or 'dose'.

When two different specific interventions are compared, a standard design is to test a new intervention against another (comparator) intervention which has previously been shown to be effective. In this case, if there is no statistically significant difference between the two conditions at the end of the trial, there is a tendency to conclude that the new intervention is not effective, when

actually it just means that it is no more (or less) effective than the comparator intervention. In fact, this kind of trial can only provide information on comparative effectiveness but cannot inform on absolute effectiveness - with no difference between the interventions, it could be that both are effective or both are ineffective. In order to distinguish between these possibilities, a no-intervention control arm of the trial is required.

In summary, researchers (and consumers of research) should be aware that the control and intervention conditions that are contrasted within an RCT constrain the conclusions that can be drawn (see for example, Brady et al., 2018).

4) Some RCTs cannot determine predictors of improvement

It is now increasingly common for RCTs to investigate the participant characteristics that interact with (moderate) the effects of treatment. A clear description of the approach used to determine mediators and moderators of treatment outcomes in psychiatry is provided by Kraemer, Wilson, Fairburn and Agras (2002). However, there are several factors which lead to RCTs, as usually employed, being unable to determine which individual characteristics (e.g., language profile, cognitive skills) relate to the likelihood or degree of improvement as a result of intervention (i.e., are moderators).

First, and most critically is the fact that RCTs do not provide a reliable measure of individual improvement. While the extent of change from pre-test to post-test will be available for each individual, we cannot determine how far this change is due to the intervention, and how far it is due to spontaneous recovery, development or placebo effects. Moreover, it is not possible to account for the extent of these confounding factors as we cannot assume that every individual will be influenced by them to the same degree: some children's language will develop more than others during the period of the study, some adults with aphasia will show more spontaneous recovery than

others and recent research demonstrates that this can continue more than a year post stroke (Hope et al., 2017). Hence for any one individual who shows an improvement of X%, all, some or none of this improvement may be due to the treatment and Y% due to the other factors. Consequently, any attempts to, for example, correlate extent of improvement with other factors, such as severity, semantic impairment, working memory etc., cannot provide reliable information on causality.

Second, in order for RCTs to have sufficient power to demonstrate differences between groups they generally include a large number of participants. Considerable resources are necessary to ensure randomisation and blinding, and care is needed in selecting primary outcome measures. Because of this, particularly the need for sufficient numbers, detailed assessment is not usually feasible. Hence, the possibility of linking outcome with background profile is often limited by a lack of detailed data.

However, there are RCTs which, while not being able to determine individual factors affecting improvement, can nonetheless inform mechanisms underlying the treatment outcomes (also termed mediators, Kraemer et al., 2002). For example, Evy Visch-Brink and colleagues in the Netherlands are carrying out a series of trials working with adults with aphasia (labelled the RATS trials). In Study One 'BOX' (a semantic approach to intervention) was compared with 'FIKS' (a phonological approach) (Doesburgh et al., 2004). While there was no significant difference in improvement between the groups receiving the different interventions overall, there were changes on specific tasks that linked with the intervention approach: the group working with FIKS improved more on phonological measures (repetition of nonwords, auditory lexical decision), while those working with BOX improved more on a semantic measure (semantic association test). Thus, the area targeted in intervention influenced this aspect of language processing, thereby providing strong evidence that the improvement from each intervention was underpinned by different mechanisms. In this way, the study helps us move from the broad question 'does therapy work?' towards answering 'which specific intervention works, for whom and how?'

While there is more research on interventions with adults with aphasia than in children with developmental language disorder, even in this field there is still a need to increase the evidence base for best practice by investigating the relationship between the nature of an individuals' aphasia and intervention outcome. Meta-analyses of aphasia therapy studies in the 1990s (Robey, 1994 and 1998) drew two main conclusions: (a) Considerable evidence has accumulated that treatment, generally considered, is effective; this trend continues, for example:

> Among the three Class I (RCT) studies and four Class II studies, comparing language remediation with no treatment, six studies with 676 subjects report significant benefits of language remediation and one class II study with 38 subjects reported no clear effect. (Cicerone et al., 2000, p. 1604)

and (b) further studies to reinforce the general conclusion would waste resources required to test more focused hypotheses. There are also repeated calls for research to evaluate the effects of cognitive rehabilitation on relevant, functional outcomes (such as everyday conversation).

Finally, RCTs using multicomponent interventions, can limit the possibility of determining which aspects of the intervention are the 'active' ingredients. For example, an important large RCT with children with developmental language difficulties (Boyle, McCartney, Forbes, & O'Hare, 2007) used a manualised therapy. The manual listed activities to support, for example, comprehension monitoring and the development of vocabulary, grammar or narrative. Children in the trial had individual learning goals and intervention was selected from the manual. The study found significant improvement in expressive language compared to a 'usual care' group. This RCT is important for justification of SLT/P service provision. However, as the intervention was complex and each child received different combinations of the elements from the manualised treatment, the study does not inform understanding of the effectiveness of different specific sub-interventions. As Boyle et al.,

(2007, p. 98) acknowledge, while important, RCTs can nevertheless leave us with the need for further specific research to answer the question of which intervention components are effective for which communication profiles.

However, there are new research designs that aim to address this problem. For example, Sequential, Multiple Assignment, Randomised Trials (SMARTs) entail adaptive intervention with a series of stages. Over multiple stages, participants are randomly assigned to a treatment option, enabling the evaluation of the effectiveness of each treatment (e.g., Lei, Nahum-Shani, Lynch, Oslin, & Murphy, 2012; for another example see also factorial designs, e.g., Collins, Dziak, Kugler, & Trail, 2014). By examining mediators, i.e., variables that interact with treatment outcome (e.g., Clarke, Snowling, Truelove & Hulme, 2010), or using a more complex design (such as SMARTs), researchers may be able to identify 'active' therapeutic components (e.g. tasks, spacing of treatment sessions) which may be further developed, and redundant elements which could be discarded.

5) Other issues

As outlined in detail by Hahs-Vaughn and Nye (2008), a major issue for RCTs is the occurrence of factors that may jeopardise quality randomisation. We draw particular attention to two of these, attrition and sample size.

Attrition: The quality of randomisation, and comparability of groups, may be compromised by attrition bias. This refers to participants being more likely to drop out of one group than the other. It could be that those who are receiving a control (placebo) intervention may be less likely to complete all testing sessions, or those receiving a highly intensive treatment may drop out due to its onerous nature. Nevertheless, attrition is well recognised and there are several strategies available for its reduction (Shadish et al., 2002). For example, placing all participants in a placebo group prior to

randomisation may reduce attrition bias, as most attrition may then occur before random assignment (Hahs-Vaughn & Nye, 2008).

Sample size: RCTs are resource intensive and therefore expensive (e.g., Sanson-Fisher, Bonevski, Green, & D'Este, 2007). To ensure sufficient power to detect an effect, relatively large numbers of participants are required, with screening of many more to ensure they meet the inclusion criteria. In some areas, where a condition is rare (e.g., primary progressive aphasia), it may simply not be possible to recruit sufficient numbers of participants for a fully powered study, within a feasible time span. Furthermore, studies with multiple arms (often useful in aiding understanding of what is effecting change) can particularly increase the sample size. The large sample sizes that are typically required to ensure that there is sufficient statistical power to detect an effect results in costly research. Similarly, the more detailed the assessment, and hence the better characterised the individuals within a study, the greater the expense.

Single case experimental designs and case series.

Observational versus experimental case studies

First let us be clear we are not discussing observational case studies in this paper. Observational studies (also known as clinical descriptions or case reports) describe individual cases in detail, but do not systematically manipulate treatment conditions. They also often do not investigate or control for stability of performance over time or control for other factors that might influence outcome (Perdices & Tate, 2009). Observational studies provide clinically useful information describing individuals' symptoms and may give some indication of the outcome of interventions. They are usually held to be important for rare diseases, can be very influential and have an illustrious history (particularly in psychiatry and neuropsychology). However, such studies generally cannot

unequivocally attribute changes in symptoms to intervention. In this paper, we explore instead the value of experimental case studies or series of such studies (case series). These are now commonly referred to as SCEDs: Single Case Experimental Designs[2]. The key difference is that, in experimental studies, interventions are evaluated scientifically using established methods to ensure that changes in symptoms can indeed be attributed to the intervention rather than any other cause (e.g., Byng & Coltheart, 1986; Franklin, 1997; Howard, Best & Nickels, 2015; Nickels, Best and Howard, 2015; Rvachew & Matthews, 2017; Perdices & Tate, 2009; Tate et al., 2008, 2013).

Observational and experimental single case studies are often confused and the boundary between them is not always clear. The medical viewpoint remains that case studies are 'low hanging fruit' for the clinician (Rothwell, 2010) and the potential value of SCEDs and the experimental control that can be employed to excellent use within these studies has until recently remained unrecognised. This is beginning to change, and in 2011 the Oxford Centre for Evidence Based Medicine placed a kind of single case design at the top of the hierarchy of evidence for the effectiveness of an intervention in an individual alongside systematic reviews and meta-analyses of RCTs (Oxford Centre for Evidence Based Medicine, 2011). These *n*-of-1 randomised controlled trials require a random sequence of several intervention and placebo phases in an individual patient with intervening washout phases. In these studies, both patient and clinician should be blind to the phases and the effect of the intervention/placebo on the behaviour of interest (say blood pressure) are monitored (see, e.g., Guyatt, Sackett, Taylor, Chong, Roberts, and Pugsley's (1986) classic double-blind *n*-of-1 randomised trial, which investigated the efficacy two asthma drugs). Hence, multiple demonstrations of a better outcome with the active intervention compared to the placebo provide evidence for the

---

[2] Although originally used to refer to one specific approach to maintaining experimental control within a single case intervention study (e.g., McReynolds & Kearns, 1983) the term is now being used more broadly to encompass other designs which also maintain experimental control (e.g., Howard et al., 2015; Nickels et al., 2015).

effectiveness of the intervention for an individual patient. Findings from *n*-of-1 RCTs are therefore considered equally strong evidence that a treatment is effective as recommendations from the meta-analysis of RCTs[3]. The full double-blind, washout *n*-of-1 design may not be feasible for all medical interventions (e.g., antibiotic prescription where a successful treatment will result in the disease process being halted) nor most speech-language interventions (where both blinding and washout are problematic). Nevertheless, in the same way as standard group-based RCTs have different designs to cope with different treatment types, the same can be true of an *n*-of-1 RCT. Many single case experimental designs can be considered as *n*-of-1 RCTs.

Key Features of Single Case Experimental Designs

Single case experimental designs are experimentally controlled designs where the researcher controls intervention and no-intervention phases and monitors change (e.g., Byng & Coltheart, 1986; Franklin, 1997; Kazdin, 1982; Thompson, 2006). Rather than employing a group of control participants to compare to an intervention group, each participant acts as their own control. Randomisation should occur but this is of items, conditions or interventions/intervention phases rather than of participants into different groups (see Nickels et al., 2015 for a discussion of randomisation within SCEDs). There are many different designs that can be employed which fall within the category of SCEDs. These include for example, crossover designs, multiple baseline designs, and withdrawal/reversal designs.

Key to all of the designs is the exercise of 'experimental control' in order to demonstrate that change occurs in behaviour is as a direct result of intervention (Thompson, 2006). A number of different methods can be employed to provide this control, such as requiring a stable pre-intervention

---

[3] It is unclear to what extent the Oxford Centre for Evidence Based Medicine consider that *n*-of-1 trials are only the highest level of evidence for the particular individual with whom the trial has been conducted, and whether they would also argue (as we do here) that the trial provides high quality evidence that may be applicable to another individual with the same characteristics.

baseline, a set number of pre-intervention assessments, treating in one modality whilst measuring change across modalities, treating one aspect of language while reassessing on other language control tasks (Howard et al., 2015; Kazdin, 2011; McReynolds & Kearns, 1983; Nickels et al., 2015; Rvachew & Matthews, 2017; Logan, Hickman, Harris, & Heriza, 2008; Tate et al., 2013, 2015). This control is necessary for us to be sure it is the intervention that is effecting the change rather than practice, placebo, or any other cause. It can also enable exploration of the mechanisms by which change may be occurring when a series of experimentally controlled case series are performed across individuals using the same methods, materials and design (Thompson, 2006).

In the cognitive neuropsychological and psycholinguistic approaches to language difficulties, theories of typical language processing underlie intervention approaches. Moreover, differences between individuals are seen as key to understanding language difficulties and to informing language intervention (e.g., Chiat, Law, & Marshall, 1997). Language processing is complex (Ellis, Young, & Anderson, 1988; Kay, Lesser, & Coltheart, 1996) and, in development, subject to many influences both internal and external to language and to the child. Moreover, patterns of impairment may vary with development (Snowling, Duff, Nash, & Hulme, 2016). Consequently, the number of different possible varieties of language impairment is extremely large, as reflected in heterogeneity even within a particular language disorder. It is therefore important to be able to determine not only whether an intervention is effective for a particular individual but what the mechanism is underlying this improvement - only then will it be possible to accurately predict which treatment will be effective for a particular individual (Best & Nickels, 2000). Case series of SCEDs provide the means to do this by exploring the relationship between language processing profiles and the outcome of intervention. For example, Howard, Hickin, Redmond, Clark and Best, (2006) explored facilitation (using word-to-picture matching) of word finding in a case series of 17 adults with anomia as part of their aphasia. The participants whose word retrieval benefited most from this approach were found

to be those whose difficulty was in accessing word forms (i.e., their profiles were characterised neither by a semantic difficulty nor by a phonological difficulty). In a similar theoretically motivated study, Best et al., (2013) found only participants with anomia characterised by less semantic difficulty and more of a phonological output deficit showed generalisation to untreated items.

The strengths and limitations of SCEDs are considered below. While the literature employing this methodology is growing, there remain debates over the best experimental design (see Howard et al, 2015; Nickels et al, 2015; Smith, 2012; Tate et al., 2008, 2013) and poorly controlled studies continue to be published under the SCED umbrella, as for RCTs. Indeed, many of the important design principles for RCTs also apply to SCEDs. Take, for example, the importance of randomising to avoid bias. As noted above, in RCTs this applies to participants, and in SCEDs this can apply to items or treatment types. Similarly, the conclusions that can be drawn from both designs rely on the relationship (and differences between) the intervention and control. In RCTs this refers to control arms of the trial (e.g., whether a no treatment or a comparator control) and in SCEDs to the nature of the contrasts between the different phases (e.g., whether the baseline phase is matched with intervention phase(s) for SLT/P contact).

<u>Strengths of SCEDs</u>

1) Results apply to individuals.

A key strength of the single case approach is that we can be confident as to whether the treatment was effective for that individual - the design allows us to conclude that the treatment has caused a change in behaviour. Moreover, by looking at the individual, an assumption of homogeneity is no longer required. If, you have a bad back which improves each time you are on tablet X but not when you are on tablet Y in a SCED study, then you know tablet X is more beneficial for you than tablet Y. As noted above, this design (the '*n*-of-1' design) is increasingly acknowledged as the optimum design

for determining that an intervention is effective for an individual (e.g., Guyatt et al., 2000; Oxford Centre for Evidence-Based Medicine Working Group, 2011). Moreover, in a study involving a case series of SCEDs, we can go beyond the group average and confidently state how many of the case series benefitted individually (and statistically evaluate the strength and variability of the effect). As noted, this is in contrast to the between groups RCT design, where we cannot determine for any individual within the group whether the treatment caused any change in behaviour.

2) SCED-Case series designs allow exploration of the link between background profile and outcome

A clinically crucial question is what works for whom? In SLT/P, the clinician working with an individual needs to be able to consider a range of approaches that may be effective and does not want to offer those that are less likely to be helpful. For example, would better outcomes be obtained if work on conversation is directly with a person with agrammatic aphasia or with their main conversational partner and, critically, is the answer different for different people? The answer can be obtained through research with a variety of people with aphasia and their conversational partners and exploring which aspects of their profile predict the outcome of each kind of intervention.

In order to be able to determine the relationship between profile and response to intervention, the first prerequisite is a reliable measure of whether an intervention is effective for an individual. This must be coupled with a replication of the intervention across individuals with both similar and different profiles. The final requirement is a detailed assessment of any aspect of the individual that may influence outcome, and for this detailed assessment to be replicated across individuals using the same measures. In SLT/P research these assessments may be within the domain of language but can also extend beyond to include other cognitive skills (e.g., memory, attention), psychological

state (e.g., depression) and, increasingly, self-ratings of activity and participation (e.g., Fillingham, Sage & Lambon-Ralph, 2006; Greenwood, Grassly, Hickin & Best, 2010).

Not only can case series address 'what works for whom', the same data can be used to address *how* the intervention has had its effects (e.g., Best, Herbert, Hickin, Osborne, & Howard, 2002). Here, it is the heterogeneity that is a problem for RCTs, which is essential to examine the effect of different profiles on performance in order to build hypotheses of the mechanism underlying treatment effects.

In sum, RCTs rarely address the question of what works for whom, and we would argue that, as they are unable to determine effectiveness at the individual level, they *cannot* answer this question. In contrast, a series of SCEDs provides a natural vehicle for such investigations. Consequently, Howard (1986) and Hegde (2007) have argued that well-designed SCEDs may provide clearer evidence on what works and for whom, and that SCED-Case series are particularly appropriate (Howard, 2003; Nickels, 2002).

3) Proximity to clinical practice.

In order to explore our intuitions that SCEDs provided evidence that was more easily applicable to clinical practice, we conducted an online survey. 144 UK speech and language therapists working with either children with specific language needs or adults with acquired aphasia responded (see the Appendix for further details of the method and the full questionnaires). The key areas investigated were the SLTs' awareness of studies (RCTs and SCEDs) that have been carried out to investigate the effectiveness of SLT/P intervention for the relevant client group, and how these two different types of research influenced their management and clinical practice. Results were analysed both quantitatively and qualitatively, but only selected relevant and representative aspects of these results are reported here.

89% of respondents were aware of either RCTs or SCEDs relevant to their field of intervention. A similar proportion of therapists agreed that they used RCTs (29%) and SCEDs (28%) to justify their service to commissioners. However, there was a large difference in terms of their use of the approaches to guide their clinical practice: around half (52%) agreed that they used RCTs to guide management of individual cases and 40% to plan specific therapy with clients; in contrast, more clinicians reported using SCEDs. This was both true for guiding management (85%; Chi-square (1) = 33.99, *p* <0.001) and for planning specific therapy (82%; Chi-square (1) = 50.86, *p* <0.001). Thus, while both these approaches were valued by clinicians, the survey results suggest SCEDs were significantly more widely used to inform the specifics of intervention.

With regard to comments on the two different research designs, different aspects of each were valued. Some respondents were more positive about large-scale group studies, with others highlighting the considerable impact of single case studies and case series on their clinical practice and with regard to planning future studies (see examples 5 and 6). Illustrative examples are provided below:

1.   *'The bigger the study numbers the more likely it is to influence what I do'*

2.   *'Control trials that have bigger participant numbers are more useful in terms of strength and use with commissioners. Also it provides more information on the type of service that we deliver. Case studies are more useful in terms of more specific therapy interventions'*

3.   *'I do find single case studies a really helpful clinical resource-- I think a move to case series design might be more pragmatic and widen application of findings but I am not convinced that RCT design presents the best way forward for measuring efficacy of therapy when aphasic patients are so diverse and therapy is tailored to specific difficulties'*

4.	*'Single case studies, either published or presented during presentations, seminars and conferences tend to be enthusiastic, encouraging and presented in practical detail. A child on my caseload comes to mind and I am keen to try a specific line of therapy'.*

5.	*'Many more single case studies, but with replication across different stages of the rehab 'journey'. RCTs are more highly thought of in terms of commissioning etc., but are not really suitable for this population - this needs acknowledging more widely.*

6.	*'I think SLTs should take more confidence in the single case study design as a useful research tool. I find this research evidence more informative and it translates more readily to clinical practice'.*

In sum, research employing RCTs was felt to be useful for obtaining funding and influencing commissioners of services. The results of SCEDs and case series were considered to be undervalued, closer to standard clinical practice and sometimes more easily applicable to a specific individual on a caseload than the findings from an RCT.

Limitations of SCEDs and SCED-Case series designs

1) Limited generalisability

A fundamental difficulty with individual SCEDs is the perceived lack of 'generalisability' of findings. In RCTs the findings from the group are considered to be generalisable to other individuals meeting the same criteria (but see the concerns raised above) whereas the results from SCEDS are thought to apply only to those individuals involved in that particular study. The logic behind this is that the RCT contains a 'sample' of the population and the outcome can be applied to the wider population from which this sample is drawn (the basis of parametric statistics). The SCED, in contrast, involves only a selected participant who may not be representative of the wider population.

However, when there is a case series of SCEDS, where all of the participants are selected from the

same defined population, then is there any reason that the results are any less generalisable than

those from RCTs? Moreover, many case series include the same number of participants as RCTs. For

example, Pulvermüller and Berthier (2008) in their RCT of constraint induced aphasia therapy include

ten participants in the treatment and seven in the control group, while Best et al., (2013) report 16

participants in a case series examining word retrieval treatment.

The question also arises regarding how many is enough? Supposing that 'anomia', a very common

symptom in aphasia, has been shown to be helped by interventions involving cues (e.g., Best et al.,

2002; Bruce and Howard, 1988; Laganaro, Di Pietro, & Schnider, 2003) and that a series of

individuals show significant benefit from this approach in well-designed SCEDs, at what point do we

have enough evidence to claim this is a useful approach for the majority of adults with anomia, in

the absence of a RCT? In terms of clinical practice, this question may be best addressed by

considering what is important for the specific individual, and whether a particular intervention will

be helpful for him/her. In terms of research, a possible solution is to pool the relevant SCEDs

together in a systematic review, and if applicable, a meta-analysis. To do so, we may require

evidence of effect sizes for each individual (but also see de Aguiar, Bastiaanse, & Miceli, 2016).

2) Inconsistent use of statistical analysis

Those employing SCED and case series designs differ in whether they employ statistical analyses to

analyse their findings and on the nature of the statistical tests used. Some studies use the many

probe scores contributed from the different phases to plot line graphs of the findings which are then

evaluated using visual inspection (e.g., Kiran & Thompson, 2003). While many such studies are well

regarded in the field and influential in both research and practice, the use of 'visual analysis' is

increasingly called into question (e.g., Smith, 2012, p. 521). Other studies use non-parametric

statistics (e.g., Lorenz & Ziegler, 2009; Routhier, Bier & Macoir, 2016) or simply calculate effect sizes with no further statistical analyses (e.g., Off, Griffin, Spencer, & Rogers, 2016; see Howard et al. 2015 for critique of some common methods for determining effect size). In setting out a scale for evaluation of the adequacy of SCED design, Tate et al. (2008) suggest that those employing these designs in the future should use statistical analysis to evaluate the outcomes. However, in their refinement of this rating scale, Tate el al. (2015) award full marks for analysis not only to those studies that use statistical tests (with justification), but also if there is systematic visual analysis using, for example, steps outlined by Kratchowill et al. (2010, 2013). Howard et al. (2015), however, note the weaknesses of visual analysis, and, like others, recommend statistical analysis be used. They also discuss strengths and weaknesses of some statistical techniques and provide a potential method (see also Laganaro, 2015)[4]. The variability in, and controversy regarding, the use of statistics does not help increase the acceptance of SCEDs in the wider research arena.

3) The need for large item sets

If we accept that statistical analysis is vital for interpretation of the results of SCEDs, then studies must be sufficiently powered to enable detection of significant improvement and for comparison of conditions. This requires relatively large samples of behaviour. Howard et al. (2015) note that larger item sets increase confidence in the results. They give an example to demonstrate this, noting that

---

[4] In SCED-Case series there is an additional statistical requirement to examine the variability across participants that is rarely adhered to (but see for example Best, 2005). While statistical analysis of each participant's results may show that some participants show significant effects and others do not, it is vital to examine whether there is statistical evidence for variability across participants. For example, Howard (2003) reanalyses data from a SCED-Case series by Pring, Hamilton, Harwood and Macbride (1993) and using a homogeneity test (see Leach, 1979) determines that there is statistical evidence that the participants with aphasia show different treatment effects (i.e., the effects of treatment are non-homogeneous). Another approach to determining whether there are significant differences in the effects of treatment across participants can be found within mixed effects modelling. In this approach, one can compare models with and without random slopes for participants - if the model that includes random slopes for participants has a better fit of the data than the model without then this indicates that there is evidence that participants show different effects of treatment.

on an 8-item set of stimuli, the 95% confidence limits on a 50% accuracy score range from 16% to

84%, but on a 30-item set the range is reduced to 31-69%. Thus, it is vital that the largest item sets

feasible are employed.

Discussion

RCTs and SCEDS have different design requirements, each with strengths and weaknesses. They also

have strengths and weaknesses in their applicability to SLT/P. It is important to reiterate that within

both types of research designs there can be studies of good or poor quality (see Pring, 2004, p. 295).

For example, both types of studies may fail to control for the possibility of researcher or selection

bias influencing the results. Table 1, below, provides a comparison between RCTs and SCED designs

for SLT/P intervention research and summarises the points discussed above. While we have

concentrated on the differences between the two designs, there are areas where advocates of both

approaches agree. First, is that there needs to be a move towards reporting of interventions that

enables the reader to be able to critically evaluate and replicate the design and methods of the

intervention. The CONSORT statements for RCTs (Schultz, Altman, & Moher, 2010), their extensions

to *n*-of-1 trials (CENT: Shamseer et al., 2015; Vohra et al., 2015) and similar reporting guidelines for

single case reporting of behavioural interventions (SCRIBE: Tate et al., 2015, 2016) aim to ensure

clarity of design reporting. In addition, the Template for Intervention Description and Replication

(TIDieR; Hoffman et al., 2014) provides a 12-item checklist for the reporting of interventions (brief

name, why, what (materials), what (procedure), who provided, how, where, when and how much,

tailoring, modifications, how well (planned), how well (actual)), which should improve the reporting

of the intervention methods and increase replicability. These guidelines also stress the importance

of measuring and reporting treatment fidelity - we can only be confident about the effects of a

treatment if we know that a participant, or group of participants, actually undertook the treatment

as described. In addition, for both approaches, outcome measures need to be chosen that are

reliable (e.g., show test-retest reliability) and sensitive to the (potential) effects of the treatment

(e.g., for a naming treatment, a measure of naming). Finally, it is vital to keep in mind that an

outcome may be statistically significant but not clinically and/or educationally important.

Research design issue	Randomised Controlled Trials - group RCT	Experimentally controlled case series/single case designs - SCED
Public and scientific understanding of design	Widely accepted as 'gold standard' in intervention and reviews of evidence base for practice; Well established designs.	Not well understood; often referred to as lacking control; Range of different designs employed.
Applicability to individual participants	If a difference between groups is found, this is in the group average; the intervention is not necessarily beneficial for all those that are treated.	The findings are applicable on a case by case basis and are analysed for each participant separately.
Experimental control	Established through a control group. However, the proximity of the intervention and control conditions will influence the conclusions that can be drawn.	Established in a number of ways, e.g., baseline testing, control items, modalities or tasks. Each participant acts as their own control.
Comparison between interventions	Comparison of effectiveness is usually between groups. E.g., intervention group 1, intervention group 2 and control group.	Comparisons of effectiveness within an individual or across a case series, allowing conclusions about the more appropriate intervention for an individual.
Generalisation	Findings are argued to be generalisable to others who meet the entry criteria for the study, but as the findings are based on group means, they may not generalise to the outcomes for individuals.	Findings apply to those included in the study. Replication can be used to extend the results to others with similar patterns of difficulty. Case series enable exploration of factors influencing outcome.
Heterogeneity amongst participants	As there is variability inherent in communication disorders, identifying relevant variables and matching groups is problematic. Heterogeneity in participants and outcomes results in the need for large samples.	Variability inherent in communication disorders and in intervention outcomes can be exploited to allow conclusions linking outcome to the nature of individual profiles
Random assignment	Participants are randomly assigned to the different groups. This is a requirement for a good RCT. The process by which it is done in health service trials can be costly.	Randomisation is important and occurs in a number of ways, e.g., items may be randomly assigned to intervention and control conditions, participants may be randomly assigned to intervention A, then B and vice versa.
Selective reporting & Publication Bias	Some variability, although negative outcomes have been reported.	Potential for extreme vulnerability from small scale studies. Null findings for an approach are most often reported in the context of a comparison between approaches.
Cost & Scale	Large scale studies are often necessary to provide adequate power. This is a multistage process (feasibility, pilot, phases I - IV) and often, financially costly and resource-intensive.	Clinically useful answers may be obtained from small scale studies. Large scale case series and replications may be more expensive.
Detailed participant profiling	Often not possible due to large group sizes and cost.	Frequent use of in depth assessment to profile participants language needs and strengths.

Table 1: Comparison of RCT and SCED-Case series

The most clinically informative research will enable us to answer the question: "What works for whom". To achieve this goal, we need to understand the influence of participant and treatment factors on individual responses to intervention. As discussed, many RCTs cannot answer this question as the findings reflect the group average and do not necessarily reflect the response of each of the individuals within the group.

In order to determine the relationship between language profile and outcome, the results for each individual need to be clear. The standard RCT design uses between groups comparisons for experimental control. These require only one pre-treatment and one post treatment measure but do not provide experimental control at the individual level.

We therefore suggest SCED-Case series research design is an entirely appropriate method to evaluate the effectiveness of treatment and provide clear guidance for clinicians. By evaluating the effectiveness of treatment for a series of individuals, the extent to which the effects are different across those individuals, and determining the cause of any variability that there might be, we can truly move our clinical understanding forward. In particular, SCED-Case series research should *not* simply be viewed as a stepping stone on the way to the definitive truth of an RCT, but rather, when properly designed and executed, it should be considered a crucial source of evidence in its own right, that should stand at the top of the pyramid of evidence (see table 1, Guyatt et al., 2000).

At the same time, we recognise the influence of the medical hierarchy of evidence, the value placed on RCTs for policy decisions, and the fact that there is a widespread belief that RCTs are the gold standard for evidence. Hegde (2007) notes:

> Researchers select experimental designs based on their training, experience, expertise, and research philosophy. And it will continue to be that way. Those who typically use a particular strategy will retain a healthy critical disposition toward the one they do not use. This is good

for the science ... because the sceptics of any approach will help keep the enthusiasts a

notch below extremists. (p. 30)

Funders of research and commissioners of clinical services predominantly hold the view of the

supremacy of the RCT and are unlikely to change in the near future, despite growing awareness of

the limitations of RCTs and the value of *N*-of-1 designs (Stirling, 2017). Consequently, we propose

that the way forward may be a hybrid design incorporating a randomised control trial within a

multiple baseline SCED-Case series (an example is Design 9, see Figure 7, in Ebbels, 2017) which is

analysed both at the individual level and as a group. There are now a small number of studies that

take this approach (e.g. Best et al., 2015; Smith-Lock et al. 2015).

We will illustrate this design with an example from Smith-Lock et al. (2015) who report an expressive

grammar intervention for children with developmental language disorder. The study was designed

with two pre-treatment assessments, and one group of children randomly assigned to a grammar

intervention (n = 19) and the other to a control intervention (language comprehension programme,

n = 15). The primary outcome measure was performance on a grammar elicitation task for the

treated grammatical targets. The results were analysed both at the group and individual level using

performance across all three time points. At the group level, there was a significant change in

performance on the grammar elicitation test for children who received the intervention, but not for

the children who received the control intervention. So, the results of the RCT show the intervention

to be effective at the group level. However, despite the large effect size for the group (Cohen's d =

1.24), in the individual analysis, there was no evidence for significant treatment-related

improvement for nine of the 19 children. The remaining ten children showed treatment-related

improvement on the treated grammatical targets: that is, they demonstrated no significant change

between the two pre-treatment baseline assessments, but improved significantly from the second

(and last) pre-test to the post-treatment. Moreover, additional experimental control was obtained

from assessment on elicitation of *untreated* grammatical targets - no child showed significant treatment-related improvement on these targets, indicating that their improvements were specific to the treatment rather than non-specific placebo effects.

Importantly, without the two pre-treatment baseline assessments, at least two further individuals would have appeared to show treatment effects: Participants 4 and 11 showed improved performance between the two pre-tests as well as following treatment. The inclusion of the two pre-treatment baselines enabled Smith-Lock et al. (2013) to determine in the case of these two participants that any gains following treatment were likely to be the result of development or test-retest practice rather than the treatment itself.

Finally, Smith-Lock et al. (2015) examined the profiles of the children who showed significant treatment-related improvement and those who did not to determine the factors underlying the different profiles. It was clear that the children's articulation ability was related to the effectiveness of treatment: None of the five children who failed an articulation screener (that specifically examined ability to produce the articulatory targets required to realise the grammatical suffixes being trained) showed significant treatment-related improvement. This analysis is only possible because of the clear identification of which individuals improved as a result of treatment.

Before concluding our argument for research in the field of speech and language therapy/pathology to swim against the tide of RCTs as the definitive test of intervention, we wish to be clear what we are _not_ suggesting. We do not intend that we should stop carrying out any RCTs in SLT/P research. RCTs have an important place and they certainly carry weight with policy makers and funders worldwide. Our aim rather is to question the current widespread acceptance of the medical hierarchy with RCTs seen as providing the gold standard and definitive answers, with the implicit assumption that these apply to each person treated, particularly in view of the resources required.

In addition, we wish to encourage researchers, clinicians, policy makers and funders to value other approaches, particularly experimentally controlled SCED-case series. Furthermore, for the future of the profession, and to benefit adults and children with communication disability, more emphasis needs to be placed on clients, clinicians and researchers working together to devise research questions and carry out clinically informed (not merely translational) research. When the tide turns and the problems with RCTs are more widely understood and acknowledged, we need to be standing firmly on the sand with evidence for a range of intervention approaches and shared understanding of how change can occur. In our profession as we move towards research maturity, we need to continue to be robust in putting clients' and families' interests first and embracing the complexity this involves. We need to further develop a strong portfolio of research that can be used to answer different questions to inform both service planning and therapy with individuals with speech and language needs.

References

Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gotzsche, P. C., & Lang, T. (2001). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine*, *134*(8), 663-694. doi: 10.7326/0003-4819-134-8-200104170-00012

Best, W. (2005). Investigation of a new intervention for children with word-finding problems. *International Journal of Language & Communication Disorders*, *40*(3), 279-318. doi: 10.1080/13682820410001734154

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J., & Howard, D. (2013). Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex*, *49*(9), 2345-2357. doi: 10.1016/j.cortex.2013.01.005

Best, W., Herbert, R., Hickin, J., Osborne, F., & Howard, D. (2002). Phonological and orthographic facilitation of word-retrieval in aphasia: Immediate and delayed effects. *Aphasiology*, *16*(1-2), 151-168. doi: 10.1080/02687040143000483

Best, W., Hughes, L., Kapikian, A., Shobbrook, K., Thomas, M., Roncoli, A., Fedor, A., Masterson, J., (2015, July). *Intervention for children with word- finding difficulties: A randomised control trial and case series study*. Paper presented at The Child Language Symposium, University of Warwick, UK.

Best, W., & Nickels, L. (2000). From theory to therapy in aphasia: Where are we now and where to next? *Neuropsychological Rehabilitation*, *10*(3), 231-247. doi: 10.1080/096020100389147

Bishop, D. V. M., Adams, C. V., & Rosen, S. (2006). Resistance of grammatical impairment to computerized comprehension training in children with specific and non-specific language impairments. *International Journal of Language & Communication Disorders*, *41*(1), 19-40. doi: 10.1080/13682820500144000

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T., & the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068-1080. doi: 10.1111/jcpp.12721

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., & Ravaud, P. (2008). Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Annals of Internal Medicine*, *148*(4), 295-309. doi: 10.7326/0003-4819-148-4-200802190-00008

Bowen, A., Hesketh, A., Patchick, E., Young, A., Davies, L., Vail, A., Long, A., Watkins, C., Wilkinson, M., Peral, G., Lambon Ralph, M., & Tyrrell, P. (2012). Clinical effectiveness, cost-effectiveness and service users' perceptions of early, well-resourced communication therapy following a stroke: a randomised controlled trial (the ACT NoW Study). *Health Technology Assessment, 16*(26), 1-160. doi: 10.3310/hta16260

Boyle, J., McCartney, E., Forbes, J., & O'Hare, A. (2007). A randomised controlled trial and economic evaluation of direct versus indirect and individual versus group modes of speech and language therapy for children with primary language impairment. *Health Technology Assessment, 11*(25), 1-158. doi: 10.3310/hta11250

Bruce, C., & Howard, D. (1988). Why don't Broca's aphasics cue themselves? An investigation of phonemic cueing and tip of the tongue information. *Neuropsychologia*, *26*(2), 253-264. doi: 10.1016/0028-3932(88)90078-4

Byng, S., & Coltheart, M. (1986). Aphasia therapy research: Methodological requirements and illustrative results. In E. Hjelmquist, & L.-G. Nilsson (Eds.), *Advances in psychology* (Vol. 34, pp. 191-213). Amsterdam: North-Holland.

Cicerone, K. D. (2005). Methodological issues in evaluating the effectiveness of cognitive rehabilitation. In P. W. Halligan & D. T. Wade (Eds.), *The effectiveness of rehabilitation for cognitive deficits* (pp. 43-58)*.* New York: Oxford University Press.

Cicerone, K. D., Dahlberg, C., Kalmar, K., Langenbahn, D. M., Malec, J. F., Bergquist, T. F., Felicetti, T., Giacino, J. T., Preston Harley, J. Harrington, D. E., Herzog, J., Kneipp, S., Laatsch, L., & Morse, P. A. (2000). Evidence-based cognitive rehabilitation: Recommendations for clinical practice. *Archives of physical medicine and rehabilitation*, *81*(12), 1596-1615.

Chiat, S., Law J., & Marshall, J. (Eds.). (1997). *Language disorders in children and adults: Psycholinguistic approaches to therapy*. London: Whurr Publishers Ltd.

Clarke, P., Snowling, M., Truelove, E., & Hulme, C. (2010). Ameliorating Children's Reading-Comprehension Difficulties: A Randomized Controlled Trial. *Psychological Science, 21*(8), 1106-1116. doi: 10.1177/0956797610375449

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Cohen, A. M., Stavri, P. Z., & Hersh, W. R. (2004). A categorization and analysis of the criticisms of evidence-based medicine. *International Journal of Medical Informatics*, *73*(1), 35-43. doi: 10.1016/j.ijmedinf.2003.11.002

de Aguiar, V., Bastiaanse, R., & Miceli, G. (2016). Improving production of treated and untreated verbs in aphasia: A meta-analysis. Frontiers in Human Neuroscience, 10, 468. doi: 10.3389/fnhum.2016.00468

Dechartres, A., Trinquart, L., Atal, I., Moher, D., Dickersin, K., Boutron, I., Perrodeau, E., Altman, D. G., & Ravaud, P. (2017). Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ*, *357*. doi: 10.1136/bmj.j2490

Brady, M. C., Godwin, J., Kelly, H., Enderby, P., Elders, A., & Campbell, P. (2018). Attention control comparisons with SLT for people with aphasia following stroke: methodological concerns raised following a systematic review. *Clinical Rehabilitation*, *32*(10), 1383-1395. doi: 10.1177/0269215518780487

Doesborgh, S. J., van de Sandt-Koenderman, M. W. E., Dippel, D. W. J., van Harskamp, F., Koudstaal, P. J., & Visch-Brink, E. G. (2004). Effects of semantic treatment on verbal communication and linguistic processing in aphasia after stroke: A randomized controlled trial. *Stroke, 35*(1), 141-146. doi: 10.1161/01.STR.0000105460.52928.A6

Dyson, H., Solity, J., Best, W., & Hulme, C. (2018). Effectiveness of a small-group vocabulary intervention programme: evidence from a regression discontinuity design. *International Journal of Language and Communication Disorders*, *53*(5), 947-958. doi:10.1111/1460-6984.12404

Ebbels, S. H. (2017). Intervention research: Appraising study designs, interpreting findings and creating research in clinical practice. *International Journal of Speech-Language Pathology*, *19*(3), 218-231. doi: 10.1080/17549507.2016.1276215

Ellis, A. W., Young, A. W., & Anderson, C. (1988). Modes of word recognition in the left and right cerebral hemispheres. *Brain and language*, *35*(2), 254-273. doi: 10.1016/0093-934X(88)90111-3

Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2006). The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation*, *16*(2), 129-154. doi: 10.1080/09602010443000254

Franklin, S. (1997). Designing single case treatment studies for aphasic patients. *Neuropsychological Rehabilitation*, *7*(4), 401-418. doi: 10.1080/713755544

Glogowska, M., Roulstone, S., Enderby, P., & Peters, T. J. (2000). Randomised controlled trial of community based speech and language therapy in preschool children. *BMJ*, *321*(7266), 923. doi: 10.1136/bmj.321.7266.923

Gold, S. M., Enck, P., Hasselmann, H., Friede, T., Hegerl, U., Mohr, D. C., & Otte, C. (2017). Control conditions for randomised trials of behavioural interventions in psychiatry: a decision framework. *The Lancet Psychiatry*, *4*(9), 725-732. doi: 10.1016/S2215-0366(17)30153-0

Greenwood, A., Grassly, J., Hickin, J., & Best, W. (2010). Phonological and orthographic cueing therapy: A case of generalised improvement. *Aphasiology*, *24*(9), 991-1016. doi: 10.1080/02687030903168220

Guyatt, G. H., Haynes, R. B., Jaeschke, R. Z., Cook, D. J., Green, L., Naylor, C. D., Wilson, M. C., Richardson, W. S. (2000). Users' guides to the medical literature: XXV. Evidence-based medicine: Principles for applying the users' guides to patient care. *Journal of the American Medical Association, 284*(10), 1290-1296. doi: 10.1001/jama.284.10.1290

Guyatt, G., Sackett, D., Wayne Taylor, D., Chong, J., Roberts, R., & Pugsley, S. (1986). Determining optimal therapy - Randomized trials in individual patients. *The New England Journal of Medicine, 314*(14), 889-892. doi: 10.1056/NEJM198604033141406

Hahs-Vaughn, D. L. & Nye, C. (2008). Understanding high quality research designs for speech language pathology. *Evidence-Based Communication Assessment and Intervention, 2*(4), 218-224. doi: 10.1080/17489530802646323

Hegde, M. N. (2007). A methodological review of randomized clinical trials. *Communicative Disorders Review, 1*(1), 17-38.

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., ... & Lamb, S. E. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, *348*, g1687. doi: 10.1136/bmj.g1687

Horner, R. D. & Baer, D. M. (1978). Multiple-probe technique: a variation of the multiple baseline. *Journal of Applied Behavior Analysis, 11*(1), 189-196. doi: 10.1901/jaba.1978.11-189

Hope, T. M. H., Leff, A. P., Prejawa, S., Bruce, R., Haigh, Z., Lim, L., Ramsden, S., Oberhuber, M., Ludersdorfer, P., Crinion, J., Seghier, M. L., & Price, C. J. (2017). Right hemisphere structural adaptation and changing language skills years after left hemisphere stroke. *Brain, 140* (6), 1718-1728. doi: 10.1093/brain/awx086

Howard, D. (2003). Single cases, group studies and case series in aphasia therapy. In I. Papathanasiou & R. De Bleser (Eds.), *The sciences of aphasia: From therapy to theory* (pp. 245-258). UK: Emerald Group Publishing.

Howard, D. (1986). Evaluating Intervention Beyond randomised controlled trials: The case for effective case studies of the effects of treatment in aphasia. *British Journal of Disorders of Communication*, *21*(1), 89-102. doi: **10.3109/13682828609018546**

Howard, D., Best, W., & Nickels, L. (2015). Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology*, *29*(5), 526-562. doi: 10.1080/02687038.2014.985884

Howard, D., Hickin, J., Redmond, T., Clark, P., & Best, W. (2006). Re-visiting "semantic facilitation" of word retrieval for people with aphasia: Facilitation yes but semantic no. *Cortex*, *42*(6), 946-962. doi: 10.1016/S0010-9452(08)70439-8

Jones, M., Onslow, M., Packman, A., Williams, S., Ormond, T., Schwarz, I., & Gebski, V. (2005). Randomised controlled trial of the Lidcombe programme of early stuttering intervention. *BMJ*, *331*(7518), 659. doi: 10.1136/bmj.38520.451840.E0

Kay, J., Lesser, R., & Coltheart, M. (1996). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology, 10*(2), 159-180. doi: 10.1080/02687039608248403

Kazdin, A. E. (1982). Single-case experimental designs in clinical research and practice. *New Directions for Methodology of Social & Behavioral Science, 13*, 33-47.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed). USA: Oxford University Press.

Kiran, S. & Thompson, C. K. (2003). The Role of Semantic Complexity in Treatment of Naming DeficitsTraining Semantic Categories in Fluent Aphasia by Controlling Exemplar Typicality. *Journal of Speech, Language, and Hearing Research*, *46*(4), 773-787. doi: 10.1044/1092-4388(2003/061).

Klee, T., Stringer, H., & Howard, D. (2009). Teaching evidence-based practice to speech and language therapy students in the United Kingdom. *Evidence-based Communication Assessment and Intervention*, *3*(4), 195-207. doi: 10.1080/17489530903399103

Kraemer H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, *59*(10), 877-883. doi:10.1001/archpsyc.59.10.877

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved July 1, 2018, from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26-38. doi: 10.1177/0741932512452794

Leach, C. (1979). *Introduction to statistics: A nonparametric approach for the social sciences*. New York: John Wiley & Sons.

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., & Murphy, S. A. (2012). A "SMART" Design for Building Individualized Treatment Sequences. *Annual Review of Clinical Psychology, 8*(1), 21-48. doi: 10.1146/annurev-clinpsy-032511-143152

Laganaro, M. (2015). Optimising the design of intervention studies: Should we test whether people require treatment or which intervention has the best outcome? *Aphasiology*, *29*(5), 605-608. doi: 10.1080/02687038.2014.986966

Laganaro, M., Di Pietro, M., & Schnider, A. (2003). Computerised treatment of anomia in chronic and acute aphasia: An exploratory study. *Aphasiology*, *17*(8), 709-721. doi: 10.1080/02687030344000193

Logan, L. R., Hickman, R. R., Harris, S. R., & Heriza, C. B. (2008). Single-subject research design: recommendations for levels of evidence and quality rating. *Developmental medicine & child neurology*, *50*(2), 99-103. doi: 10.1111/j.1469-8749.2007.02005.x

Lorenz, A., & Ziegler, W. (2009). Semantic vs. word-form specific techniques in anomia treatment: A multiple single-case study. *Journal of Neurolinguistics*, *22*(6), 515-537. doi: 10.1016/j.jneuroling.2009.05.003

McArthur, G., & Castles, A. (2017). Helping children with reading difficulties: some things we have learned so far. *npj Science of Learning*, *2*(1), 7. doi: 10.1038/s41539-017-0008-3

McReynolds, L. V., & Kearns, K. P. (1983). *Single-subject experimental designs in communicative disorders*. Baltimore, MD: University Park Press.

Moher, D, Hopewell, S, Schulz, K. F., Montori, V., Gotzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., Altman, D. G. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*, c332. doi: 10.1136/bmj.c332

Morgan, A. T., & Vogel, A. P. (2009). A Cochrane review of treatment for childhood apraxia of speech. *European Journal of Physical and Rehabilitation Medicine*, *45*(1), 103-110.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, *16*(10-11), 935-979. doi: 10.1080/02687030244000563

Nickels, L., Best, W., & Howard, D. (2015). Optimising the ingredients for evaluation of the effects of intervention. *Aphasiology*, *29*(5), 619-643. doi: 10.1080/02687038.2014.1000613

Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence Working Group. (2011). The Oxford Levels of Evidence 2. Retrieved July 3, 2018 from https://www.cebm.net/index.aspx?o=5653

Off, C. A., Griffin, J. R., Spencer, K. A., & Rogers, M. A. (2016). The impact of dose on naming accuracy with persons with aphasia. *Aphasiology*, *30*(9), 983-1011. doi: 10.1080/02687038.2015.1100705

Perdices, M., & Tate, R. (2009). Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued? *Neuropsychological Rehabilitation*, *19*(6), 904-927. doi: 10.1080/09602010903040691

Pulvermüller, F., & Berthier, M. L. (2008). Aphasia therapy on a neuroscience basis. *Aphasiology*, *22*(6), 563-599. doi: 10.1080/02687030701612213

Pring, T. (2004). Ask a silly question: Two decades of troublesome trials. *International Journal of Language & Communication Disorders*, *39*(3), 285-302. doi: 10.1080/13682820410001681216

Pring, T., Hamilton, A., Harwood, A., & Macbride, L. (1993). Generalization of naming after picture/word matching tasks: Only items appearing in therapy benefit. *Aphasiology, 7*(4), 383-394. doi: 10.1080/02687039308249517

Robey, R. R. (1998). A meta-analysis of clinical outcomes in the treatment of aphasia. *Journal of Speech Language and Hearing Research, 41*(1), 172-187. doi: 10.1044/jslhr.4101.172

Robey, R. R. (1994). The efficacy of treatment for aphasic persons: A meta-analysis. *Brain and Language, 47*(4), 582-608. doi: 10.1006/brln.1994.1060

Rosenbek, J. C. (2016). Tyranny of the randomised clinical trial. *International Journal of Speech-Language Pathology*, *18*(3), 241-249. doi: 10.3109/17549507.2015.1126644

Rothwell, P. (2010). Clinical research in the UK: Back from the brink, but are neurologists missing the boat? *Practical Neurology, 10*(2), 64-66. doi: 10.1136/jnnp.2010.208116

Routhier, S., Bier, N., & Macoir, J. (2016). Smart tablet for smart self-administered treatment of verb anomia: Two single-case studies in aphasia. *Aphasiology, 30*(2-3), 269-289. doi: 10.1080/02687038.2014.973361

Rvachew, S., & Matthews, T. (2017). Demonstrating treatment efficacy using the single subject randomization design: A tutorial and demonstration. *Journal of Communication Disorders*, *67*, 1-13. doi: 10.1016/j.jcomdis.2017.04.003

Sackett, D L. (1995). Applying overviews and meta-analyses at the bedside. *Journal of Clinical Epidemiology, 48*(1), 61-66. doi: 10.1016/0895-4356(94)00085-5

Sanson-Fisher, R. W., Bonevski, B., Green, L. W., & D'Este, C. (2007). Limitations of the randomized controlled trial in evaluating population-based health interventions. *American Journal of Preventive Medicine*, *33*(2), 155-161. doi: 10.1016/j.amepre.2007.04.007

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine, 152*(11), 726-732. doi: 10.7326/0003-4819-152-11-201006010-00232

Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Nikles, J., Tate, R., Johnston, B. C., Zucker, D., Shadish, W. R., Kravitz, R., Guyatt, G., Altman, D. G., Moher, D., Vohra, S., & the Cent group. (2015). CONSORT extension for reporting *N*-of-1 trials (CENT): Explanation and elaboration. *BMJ*, *350*. doi: 10.1136/bmj.h1793

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods, 17*(4), 510-550. doi:10.1037/a0029312

Smith-Lock, K. M., Leitão, S., Prior, P., & Nickels, L. (2015). The effectiveness of two grammar treatment procedures for children with SLI: A randomized clinical trial. *Language, Speech, and Hearing Services in Schools*, *46*(4), 312-324. doi: 10.1044/2015_LSHSS-14-0041

Smith-Lock, K. M., Leitao, S., Lambert, L., & Nickels, L. (2013). Effective intervention for expressive grammar in children with specific language impairment. *International Journal of Language & Communication Disorders*, *48*(3), 265-282. doi: 10.1111/1460-6984.12003

Snowling, M. J., Duff, F. J., Nash, H. M., & Hulme, C. (2016). Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry*, *57*(12), 1360-1369. doi: 10.1111/jcpp.12497

Stirling, C. (2017, October 20). Move over RCT – Time for a revised approach to evidence based medicine [Blog post]. Retrieved from https://blogs.bmj.com/bmj/2017/10/20/christine-stirling-move-over-rct-time-for-a-revised-approach-to-evidence-based-medicine/

Tate, R. L., & Perdices, M. (2015). *N-of-1 Trials in the Behavioral Sciences.* In J. Nickles & G. Mitchell (Eds.), *The Essential Guide to N-of-1 Trials in Health* (pp. 19-41). doi: **10.1007/978-94-017-7200-6**

Tate, R. L., Mcdonald, S., Perdices, M., Togher, L., Schultz, R., & Savage, S. (2008). Rating the methodological quality of single-subject designs and *n*-of-1 trials: Introducing the Single-Case Experimental Design (SCED) Scale. *Neuropsychological Rehabilitation*, *18*(4), 385-401. doi: 10.1080/09602010802009201

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., Horner, R., Kazdin, A., Kratochwill, T., McDonald, S., Sampson, M., Shamseer, L., Togher, L., Albin, R., Backman, C., Douglas, J., Evans, J. J., Gast, D., Manolov, R., Mitchell, G., Nickels, L., Nikles, J., Ownsworth, T., Rose, M., Schmid, C. H., & Wilson, B. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 Statement. *Aphasiology*, *30*(7), 862-876. doi: 10.1080/02687038.2016.1178022

Tate, R. L., Perdices, M., Rosenkoetter, U., Wakim, D., Godbee, K., Togher, L., & McDonald, S. (2013). Revision of a method quality rating scale for single-case experimental designs and *n*-of-1 trials: The 15-item Risk of Bias in *N*-of-1 Trials (RoBiNT) Scale. *Neuropsychological Rehabilitation*, *23*(5), 619-638. doi: 10.1080/09602011.2013.824383

Tate, R. L., Rosenkoetter, U., Wakim, D., Sigmundsdottir, L., Doubleday, J., Togher, L., McDonald, S., & Perdices, M. (2015). *The risk-of-bias in N-of-1 trials (RoBiNT) scale: An expanded manual for the critical appraisal of single-case reports*. Sydney, Australia: John Walsh Centre for Rehabilitation Research.

Thompson, C. K. (2006). Single subject controlled experiments in aphasia: The science and the state of the science. *Journal of Communication Disorders*, *39*(4), 266-291. doi: 10.1016/j.jcomdis.2006.02.003

Varley, R., Cowell, P. E., Dyson, L., Inglis, L., Roper, A., & Whiteside, S. P. (2016). Self-administered computer therapy for apraxia of speech: two-period randomized control trial with crossover. *Stroke*, *47*(3), 822-828. doi 10.1161/STROKEAHA.115.011939

Vohra, S., Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Tate, R., Nikles, J., Zucker, D. R., Kravitz, R., Guyatt, G., Altman, D. G., Moher, D. (2015). CONSORT extension for reporting *N*-of-1 trials (CENT) 2015 statement. *BMJ, 350*, h1738. doi: 10.1136/bmj.h1738

Wallace, S. J., Worrall, L., Rose, T., & Le Dorze, G. (2016). Core outcomes in aphasia treatment research: An e-Delphi consensus study of international aphasia researchers. *American Journal of Speech-Language Pathology*, *25*(4S), S729-S742. doi: 10.1044/2016_AJSLP-15-0150

APPENDIX

Survey Method

An online questionnaire was made available via Opinio, a web-based survey tool. Participants were approached by email, via the membership secretaries of Royal College of Speech and Language Therapists Clinical Excellence Networks (CENs) for therapists working with these client groups (in total ten developmental and five acquired CENs), and the British Aphasiology Society (BAS). The membership secretaries were sent an email, containing information about the study and a link to the Opinio survey, and were requested to forward this to their membership.

The email explained the purpose of the study, and informed participants that all data would be collected anonymously with the online survey being open for one month. The email also stated that completion of the questionnaire would be taken to indicate informed consent.

Two versions of the questionnaire were produced. These were identical except for the choices provided for the clinical setting: one for SLTs working with adults with acquired aphasia, the other for SLTs working with children with specific language needs. These were worded the same, apart from references in the question to the specific client group and the clinical setting (see below).

There were 12 questions in total. The first three probed background information, question four asked respondents to indicate how often they drew on different sources of information when planning specific intervention activities for the relevant client group.

Therapists rated the extent to which they agreed with statements about how they used the results of Randomised Control Trials (RCTs – large group studies comparing the effects of intervention against a control group) and Single Case/Case series Experimental Designs (SCEDs – single cases or case series where each participant acts as their own control).

The next three questions (5-7) focused specifically on awareness of RCTs. Questions 8-10 followed

the same format to probe awareness of and use of experimental single case studies or case series.

At the end of the questionnaire there were two open questions to determine the influence of

different types of research on clinical practice and to determine future priorities

Content validity

The questionnaire went through a number of revisions to produce a draft that was circulated to five

specialist SLTs who complete the questionnaire and provided feedback and suggestions about both

the general format of the questions and the specific wordings. This feedback was used to produce

the final version of the questionnaire.

Respondents

In total, 144 SLTs completed the questionnaires: 30 (21%) were SLTs working with children with

specific language needs, 114 (79%) worked with people with acquired aphasia. There was

considerable variation in clinical experience within the group, respondents ranging from one to

forty-three years post-qualification (median = 10 years). All respondents were currently working

with the specific client group for at least one clinical session (half day) per week (median = 5.5

sessions, range 1-10). Respondents were based in a broad range of clinical settings, including

community clinics, language resource bases, early year's settings, mainstream schools and

independent practice with children; acute, rehabilitation, community and domiciliary settings with

adults. 48% of the group were based in one setting, with the others working across two (33%) or

more (19%).

Questionnaire for SLT/Ps working with adults with acquired aphasia

We are interested in finding out more about the influence on Speech and Language Therapists' choice of specific therapy activities and how SLT intervention is informed by the evidence base in the literature. The results of the questionnaire will be used as part of a journal article, which will explore the value of different types of research on your practice, whether this is directly through reading articles or via other routes such as discussion with colleagues.

(1) Please list the clinical setting(s) in which you work with adults with aphasia

__ Acute hospital
__ Rehabilitation setting
__ Community
__ Domiciliary
__ Other (please specify):

(2) How many sessions (i.e., half-days) per week do you work with adults with aphasia? ___sessions

(3) Number of years post-qualification?   ___ years

Before completing the questionnaire, please think about the types of interventions you use with adults with aphasia. This will include many elements such as frequency of contact, working with or through others and also the specific focus of therapy. Consider activities that you have used recently to address communication goals with adults with aphasia (e.g., What was the therapy goal? Where did you get the idea for these activities from?)

(4) When planning specific intervention activities for adults with aphasia on your caseload, how often do you draw on the following sources of information:

	Always	Often	Sometimes	Rarely	Never
Clinical experience					
Consultations with colleagues					
Pre-qualification training					
Post-qualification CPD courses					
SIG meetings					
Clinical practice guidelines					
Professional publications (e.g., RCSLT bulletin)					
Textbooks					
Journal articles					
Online resources					
Other (please specify)					

Randomised Control Trials
(5) Are you aware of any randomised control trials (RCTs) that have been carried out to investigate the effectiveness of SLT intervention for adults with aphasia? (RCTs are large group studies comparing the effects of intervention against a control group)

Yes ____   No   ____

(6) Please indicate the extent to which you agree with the following statements. I use the results of RCTs to

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Justify my service to commissioners					
Guide my management of individual cases					
Plan specific therapy with clients					

Any comments:

(7) If possible, please provide examples of any RCTs that have influenced your practice. Knowledge of these may be via e.g., your training, CPD, reading journals, SIGs, colleagues. (A full reference is not necessary – please name the author(s) and give an indication of general topic area and any further details).

Single Case Experimental Designs
(8) Are you aware of any experimental case studies or vase series that investigate the effectiveness of SLT intervention for adults with aphasia? (Such studies are sometimes known as SCEDs – Single Case Experimental Design(s). They involve single case studies or case series where each adult acts as their own control).

Yes ____   No   ____

(9) Please indicate the extent to which you agree with the following statements. I use the results of SCEDs to

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Justify my service to commissioners					

Guide my management of individual cases					
Plan specific therapy with clients					

Any comments:

```



```

(10) If possible, please provide examples of any single case studies or case series that have influenced your practice. Knowledge of these may be via e.g., your training, CPD, reading journals, SIGs, colleagues. (A full reference is not necessary – please name the author(s) and give an indication of general topic area and any further details).

```



```

(11) Please add any comments about how different types of research influence your clinical practice in this area.

```



```

(12) What do you think priorities should be for future research to inform SLT management and intervention for adults with acquired aphasia?

```



```

Questionnaire for SLT/Ps working with children with specific speech and language needs

We are interested in finding out more about the influence on Speech and Language Therapists' choice of specific therapy activities and how SLT intervention is informed by the evidence base in the literature. The results of the questionnaire will be used as part of a journal article, which will explore the value of different types of research on your practice, whether this is directly through reading articles or via other routes such as discussion with colleagues.

(1) Please list the clinical setting(s) in which you work with children with specific speech and language needs

__ Mainstream school
__ Community clinic
__ Language resource base
__ Other (please specify):

(2) How many sessions (i.e., half-days) per week do you work with children with specific speech and language needs?   ___sessions

(3) Number of years post-qualification?   ___years

Before completing the questionnaire, please think about the types of interventions you use with children with specific speech and language needs. This will include many elements such as frequency of contact, working with or through others and also the specific focus of therapy. Consider activities that you have used recently to address communication goals with children with specific speech and language needs (e.g., What was the therapy goal? Where did you get the idea for these activities from?)

(4) When planning specific intervention activities for children with specific speech and language needs on your caseload, how often do you draw on the following sources of information:

	Always	Often	Sometimes	Rarely	Never
Clinical experience					
Consultations with colleagues					
Pre-qualification training					
Post-qualification CPD courses					
SIG meetings					
Clinical practice guidelines					
Professional publications (e.g., RCSLT bulletin)					
Textbooks					
Journal articles					
Online resources					
Other (please specify)					

Randomised Control Trials

(5) Are you aware of any randomised control trials (RCTs) that have been carried out to investigate the effectiveness of SLT intervention for children with specific speech and language needs? (RCTs are large group studies comparing the effects of intervention against a control group)

Yes ____   No   ____

(6) Please indicate the extent to which you agree with the following statements. I use the results of RCTs to

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Justify my service to commissioners					
Guide my management of individual cases					
Plan specific therapy with clients					

Any comments:

(7) If possible, please provide examples of any RCTs that have influenced your practice. Knowledge of these may be via e.g., your training, CPD, reading journals, SIGs, colleagues. (A full reference is not necessary – please name the author(s) and give an indication of general topic area and any further details).

Single Case Experimental Designs
(8) Are you aware of any experimental case studies or vase series that investigate the effectiveness of SLT intervention for children with specific speech and language needs? (Such studies are sometimes known as SCEDs – Single Case Experimental Design(s). They involve single case studies or case series where each adult acts as their own control).

Yes ____   No   ____

(9) Please indicate the extent to which you agree with the following statements. I use the results of SCEDs to

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Justify my service to commissioners					
Guide my management of individual cases					
Plan specific therapy with clients					

Any comments:

```



```

(10) If possible, please provide examples of any single case studies or case series that have influenced your practice. Knowledge of these may be via e.g., your training, CPD, reading journals, SIGs, colleagues. (A full reference is not necessary – please name the author(s) and give an indication of general topic area and any further details).

```



```

(11) Please add any comments about how different types of research influence your clinical practice in this area.

```



```

(12) What do you think priorities should be for future research to inform SLT management and intervention for children with specific speech and language needs?

```



```