Title: **An Umbrella Review of Aphasia Intervention descriPtion In Research: the AsPIRE project**

Short title: **Quality of Aphasia Intervention Description in Research**

Authorship: L. T. Dipper*, S. Franklin, V. de Aguiar, A. Baumgaertner, M. Brady, W. Best, S. Bruehl, G. Denes, , E. Godecke, M. Gil, M. Knoph, M. Kirmess, C. Markey, M. Meinzer, C. Mendez Orellana, F. Nouwens, M. L. Rose, M. van de Sandt, A. Whitworth, E. G. Visch-Brink

Corresponding author: Dr Lucy Dipper,

Division of Language and Communication Science

School of Health Sciences

City University London

Northampton Square

London EC1V 0HB

United Kingdom

Tel: +44(0)20 7040 4658, l.t.dipper@city.ac.uk

Abstract

Background: Recent reviews conclude that aphasia intervention is effective. However, replication and implementation require detailed reporting of intervention is and a specification of participant profiles. To date, reviews concentrate more on efficacy than on intervention reporting quality.

Aims: The aim of this project is to review the descriptions of aphasia interventions and participants appearing in recent systematic reviews of aphasia intervention effectiveness. The relationship between the quality of these descriptions and the robustness of research design is explored, and the replicability of aphasia interventions is evaluated.

Methods and Procedures: The scope of our search was an analysis of the aphasia intervention studies included in the Brady et al. 2016 and EBRSR 2018 systematic reviews, and in the RCSLT 2014 literature synthesis. Intervention descriptions published separately from the intervention study (i.e. published online, in clinical tools, or a separate trial protocols) were not included. The criteria for inclusion were that participants had aphasia, the intervention involved language and/or communication, and included the following research designs: Randomised Controlled Trial (RCT), comparison or control, crossover design, case series. Exclusion criteria included non-SLT interventions, studies involving fewer than four participants, conference abstracts, studies not available in English. Studies were evaluated for completeness of intervention description using the TIDieR Checklist. Additionally, we rated the quality of patient and intervention description, with particular reference to replicability.

Outcomes and Results: Ninety-three studies were included. Only 14 studies (15%) had >50 participants. Fifty-six studies (60%) did not select participants with a specific aphasia profile, and a further 10 studies only described participants as non-fluent. Across the studies, an average of eight (of 12) TIDieR checklist items were given but information on where, tailoring, modification and fidelity items was rarely available. Studies that evaluated general aphasia intervention approaches

tended to use RCT designs, whereas more specific intervention studies were more likely to use case series designs.

Conclusions: Group *s*tudies were generally under-powered and there was a paucity of research looking at specific aphasia interventions for specific aphasia profiles. There was a trade-off between the robustness of the design and the level of specificity of the intervention described. While the TIDieR framework is a useful guide to information which should be included in an intervention study, it is insufficiently sensitive for assessing replicability. We consider possible solutions to the challenges of making large-scale trials more useful for determining effective aphasia intervention.

Background

A framework for the adequate description of aphasia interventions is required, with robust data about intervention content, intervention method, patient selection and factors including intervention intensity, frequency and duration. In 2016, Brady and Colleagues (Brady, Kelly, Godwin, Enderby & Campbell, 2016) carried out a Cochrane review of speech and language therapy (SLT) for aphasia. They assessed the effects of SLT for aphasia following stroke, from published randomised controlled trials (RCTs) in peer reviewed journals. They concluded that SLT for people with aphasia following stroke was effective in terms of improved functional communication, reading, writing, and expressive language compared with no intervention. The authors of the review note that reporting guidelines had enhanced the quality of the description in the more recent trials. However, there are still evident limitations including the reported detail that hinders replication (to strengthen the evidence base) and implementation in clinical practice. This paper sets out to review the aphasia intervention evidence base in terms of the level of detail in intervention reporting.

Guidelines for intervention reporting

A variety of frameworks are available to guide the reporting of interventions. For trial protocols, the SPIRIT statement (Standard Protocol Items: Recommendations for Interventional Trials; Chan et al., 2013) provides guidance about content. For reporting findings from randomized controlled trials, the key guidelines are the CONSORT statement (Consolidated Standards of Reporting Trials; Schulz, Altman, Moher & CONSORT Group, 2010). These statements advises that authors should report on interventions in enough detail to allow replication (SPIRIT, item 11; CONSORT, item 5) but do not further specify how this should be done. The Template for Intervention Description and Replication (TIDieR) (Hoffmann et al., 2014), extends CONSORT 2010 (item 5) and SPIRIT 2013 (item 11). This template is designed to provide a guide for the description of intervention in stroke rehabilitation (detailed in the Methods, table 1) to stipulate sufficient detail for replication studies. The benefit of

the use of the checklist in different areas of stroke rehabilitation is encouraged (e.g., Yamato et al. 2016; Campbell et al., 2018). While the TIDieR statement was developed for adequate descriptions of intervention in the context of clinical trials, Cotterill et al. (2018) show that it is also useful for guiding intervention description more broadly. There is evidence that guidelines such as these have improved the quality of reporting, particularly in randomised controlled trials (e.g. Cobo et al., 2011). However, there remains work to be done to improve the specificity of reporting intervention detail, in particular for more complex interventions such as speech and language therapy for aphasia.

Reporting complex interventions

The reporting of complex interventions requires an increased level of detail. Describing an aphasia intervention adequately requires at least information about the content of intervention along with a description of the hypothesized therapeutic activity, including the way the intervention is conveyed to the client. Important aphasia intervention characteristics include instructional or motivational setting, cueing or coaching parameters, feedback, and reinforcement. These characteristics are required for replication and implementation but also are hypothesized to contribute to the efficacy of the intervention. At present, aphasia rehabilitation lacks both a common language by which to express these complex ingredients, and a unifying theory to help identify which aspects are most important (Whyte, 2008). The Cochrane review authors (Brady, Kelly, Godwin, Enderby & Campbell., 2016) called for future research to establish what are the optimum methods, frequency, duration, and format of SLT provision for specific patient groups. Some of this work has begun, for example a review of the effects of different treatment intensities on communication interventions (Warren, Fey & Yoder, 2007). Intervention description (following the TIDieR template) of a range of interventions may contribute to the further identification of the essential aspects (the 'active ingredients') of aphasia intervention; both in general and in terms of specific methods of intervention.

Reporting different research designs

The different research designs utilized in the evidence base may influence which aphasia intervention methods are evaluated and how these methods are described. The gold standard for clinical research is the RCT, in which individuals are randomly assigned to one of two or more study arms comparing treatment vs no treatment or comparing more than one treatment (parallel RCT). A robust RCT requires sufficient statistical power to address the clinical question and this requires a large sample size. Other research designs also contribute considerably to the aphasia evidence base, including cross-over designs, case series and non-randomized group studies. Beyond the Cochrane review of RCTs of aphasia intervention (Brady et al., 2016), other systematic reviews and literature syntheses of aphasia rehabilitation have included a wider range of research designs in their evaluation of the evidence base. For example, this is the case in the Evidence-Based Review of Stroke Rehabilitation (18th edition, Aphasia and Apraxia chapter: Faltynek et al., 2018) and the RCSLT Resource Manual for commissioning and planning services for Speech, Language, and Communication Needs (SLCN), Aphasia (Enderby & Cantrell, 2014). A treatment method is often reported in different research designs within the evidence base, over the course of its development and/or replication. efficacy may be established first by means of a case-series design, followed by a group study and/or RCT, which highlights the importance of describing the method in enough detail for replication. For these reasons, we considered it important to include a range of different research designs in this review of aphasia intervention description, in order to provide a comprehensive evaluation of the reporting in the aphasia rehabilitation evidence base.

The current study

Although there have been recent improvements in aphasia intervention reporting, the evidence base is not consistent due to the issues outlined above (lack of detail in reporting guidelines; the inherent

complexity in reporting complex interventions; and the varying levels of reporting detail characteristic of different research designs). The purpose of this review is to explore the evidence base, covering a range of research designs, and to evaluate it with respect to replicability. For this purpose, an 'umbrella review' format was chosen. An umbrella review summarizes the evidence from multiple research syntheses, to provide an overall examination of a body of information that is already available (Becker, Deeks & Oxman, 2006). In contrast with systematic reviews and literature syntheses, an umbrella review does not start with an independent search of research databases but instead draws from the reference lists of other relevant reviews. The purpose is to provide an overview of other reviews, with respect to a specific question. This umbrella review aims to synthesise three recent reviews of aphasia intervention efficacy (see Methods section for details) in order to evaluate the description of interventions and participants in the aphasia evidence base. We also aim to explore the relationship between the quality of these descriptions and the type of research design.

Method

Research design

The source for the literature search was the reference lists from three recent reviews of aphasia intervention efficacy: the Cochrane Review of Speech and language therapy for aphasia following stroke (Brady et al., 2016); the Canadian Partnership for Stroke Recovery's evidence-based review of stroke rehabilitation (Aphasia and Apraxia chapter: Faltynek et al., 2018); the Royal College of Speech and Language Therapists' Resource Manual for Commissioning and Planning Services for SLCN (Aphasia: Enderby & Cantrell, 2014). Inclusion and exclusion criteria were applied in line with our research questions, and the resulting studies were analysed in relation to the quality of the treatment descriptions and their replicability.

Research questions

The research questions were as follows:

i. What types of intervention are included in the studies?
ii. What is the quality of the studies in terms of types of research design, the use of standardized tests to measure outcomes, and the numbers of participants?
iii. Are aphasia profiles specified in inclusionary criteria?
iv. What is the quality of aphasia intervention reporting?
v. How replicable are the aphasia interventions?

Measures

Two measures were used to explore the quality of reporting and replicability:

- In order to establish whether the studies provided adequate descriptions of aphasia interventions, we analysed each study using the TIDieR checklist (Hoffman et al., 2014). The checklist was initially devised to aid those designing an intervention study, although we used it post hoc. The TIDieR items are listed in Table 1.

  [insert Table 1 here]

The scoring was binary; one point was scored if information was present, and zero points if not.

- A second qualitative checklist, the Quality of Therapy Reporting (QTR), was devised to further explore the TIDierR categories 'What' and 'How'. This is a new measure devised specifically for the current review. While the TIDieR analyses allow for a score of one on these categories when *some* information about intervention is given, we were interested in the extent to which intervention was described in a way that it was replicable. That is, having read the paper, would we be confident that we would be able to carry out the intervention? We therefore devised a further analysis, again with binary judgements, but here the judgments were based on sufficient detail for replicability. The judgements were made on five indices, listed below, which were devised through consensus decision making discussions in the author team:

    1. Was the method clear?
    2. Were materials and items specified?
    3. Was service delivery clear?
    4. Were the target responses (expected from the participant) specified?
    5. Was information given about what to do when a target response is incorrect or absent?

    Three researchers from the author team carried out the QTR rating. In order to benchmark their rating, they first looked at five papers independently; came together to discuss their responses; and to reach consensus on any discrepancies. For each question above, the guiding principle whether there was enough detail reported in the paper for an experienced clinician or researcher to replicate the treatment. As an example, consider question 2. This question achieved a score if the full stimulus list of items (e.g. nouns) were specified; there was a description of how the items were chosen such that it could be could followed; or enough examples were given such that the method of selection were clear and could be replicated. Where relevant, the paper would also need to refer to the kind of stimulus materials used (e.g. picture cards, typed sentences) in enough detail to be replicated.

Study selection

The three sources yielded 315 studies for consideration. The following inclusion and exclusion criteria were applied to the studies (see Figure 1).

Inclusion criteria:

1. Studies which included four or more participants with aphasia
2. Studies which described intervention involving language or communication
3. Studies which contained a research design with a control.

Exclusion criteria

1. Studies involving fewer than four participants
2. Studies involving interventions not targeting aphasia
3. Studies which did not describe speech and language therapy (e.g. conversation partner training, intervention for cognitive difficulties such as memory, or traditional Chinese medicine)
4. Studies with no control and/or no post-test measures
5. Studies concerning drug trials and brain stimulations studies
6. Studies which did not describe research (e.g. trial protocols, conference abstracts, systematic reviews and meta analyses, or Letters to the Editor)
7. Studies which are not published in English
8. Unobtainable studies.

[insert Figure 1 here]

Data coding and analyses

The data were characterized in terms of i) the categories of intervention type, ii) the experimental design (including design type, use of standardized outcome measures, number of participants), and

iii) the participant selection information provided. Following this, two measures were used to rate iv) the quality of the treatment descriptions and v) their replicability.

i. Categorisation by intervention type was carried out once the studies were selected, so as to reduce the number of categories while being able to clearly describe each category, based on the description of intervention given in the study.

ii. Each study was categorised in terms of one of four research design types (case series, parallel group, crossover, randomised control trial) using the following operational definitions:

Case series - Compares the effect of a treatment on a series of individuals before and after an intervention but with no control group.

Parallel group - Compares the effect of two or more treatments (one of which may be a control treatment) on a group of participants.

Crossover - In this design, each participant undertakes a sequence of two or more treatments (or a waiting period + an active treatment). In this design, each participant acts as his or her own control.

Randomized Control Trial (RCT) – Compares the effect of two or more treatments (one of which may be a control treatment) on a group of participants who have been randomly allocated to a treatment group.

In addition, studies were examined to identify whether not they used a standardized outcome measure; and to identify the number of participants. A Kruskal-Wallis test was used to explore whether there was a difference between research designs in terms of the number of participants.

iii. Studies were categorised in terms of the specificity of their participant selection criteria. Studies were categorised as having specified selection criteria (and coded 'C') if they described the selection of participants on the basis of criterion scores on specific (cognitive) linguistic tests or subtests of test-batteries (not the score for the test-battery as a whole).

Studies that specified only an overall test-battery score, diagnosis by informal description of participant characteristics, or those that only described the participant profiles details after selection, were categorised as having unspecified selection criteria (and coded 'US'). In addition, those studies selecting non-fluent participants were also identified. These three categorisation processes were carried by three of the authors (SF, EB, LD). Each person individually categorised all studies, and then discussed and agreed discrepancies in order to reach a consensus.

iv. Eighteen of the authors collectively rated the 93 studies with respect to whether the TIDieR checklist items were present. Each author individually rated a random selection of studies (mean = 5 studies; range = 2-7). Two of each person's TIDieR forms (~39% of studies) were blindly scored a second time by the first author and, where there were disagreements, the first author's scores were entered into the analysis. The inter-rater reliability was very good at 87% agreement. A Kruskal-Wallis test was used to explore whether there was a difference between research designs in terms of their TIDieR scores.

v. Three of the authors (LD, SF, EVB) undertook the QTR analysis which evaluated whether sufficient detail for replicability was reported. They each individually rated thirty-one studies. The five indices in the QTR were scored 1 (yes) or 0 (no), giving a total score of 5. As a first step, two studies were each rated by the three authors and then discussed, in order to benchmark the rating for each question. The remaining studies were then divided amongst these three authors for individual ratings, and then 10% of the studies were second-rated. There was 87% inter-rater agreement. A Pearson's correlation was used to look for an association between QTR score and number of participants; and a Kruskall-Wallis test was used to explore whether there was a difference between research designs in terms of their QTR score.

Results

Applying the inclusion and exclusion criteria described in the methods section, 93 studies remained. Details of the individual studies included are given in Appendix A. Quantitative analysis of the 93 studies, looking at categories of intervention, research design and information about participant selection will be presented first, followed by analyses of the quality of information regarding the interventions.

Studies categorized by intervention type

As described in the method, categorisation of the intervention types occurred after selection of the 93 studies, and extensive discussions among authors took place until a consensus was reached. The categories were general, lexical, communication, conversation, sentence, MIT type, reading and other. Since some studies (n=18) compared different aphasia interventions, there are more than 93 interventions included in this figure (total interventions counted = 111).

The thirty-two General interventions (29%) were those not specified in detail, or where a range of interventions were described without saying which interventions were given to which individuals with aphasia. Terms like 'usual care', 'traditional therapy', 'stimulation therapy', or 'model-based therapy' were often used to describe the intervention given. Where the term 'model-based therapy' was used, it was unclear in each case which model was being referred to. Five studies referred to computer programmes but did not specify how the intervention was administered on an individual basis. In 13 of the 32 instances, a 'general therapy' was contrasted with another intervention, so the general therapy might be considered a control.

The largest category included all lexical interventions, that is, interventions to improve auditory-verbal processing at the word level (n=40, 36%). Thirty four out of the 40 interventions in this category were to improve naming. Four interventions aimed to improve comprehension and naming;

two interventions aimed to improve comprehension only. Methods used included facilitation/cueing (22 studies), semantic tasks (nine studies), phonological tasks (four studies), gesture (three studies), AmerInd signs (one study), complex hand movements (one study).

Six studies included sentence interventions (5%) and were those that elicited sentence structure. All involved a version of mapping therapy (where verbs and their arguments are produced with the support of a sentence frame (Byng, Nickels & Black, 1994). One of these studies described VNEST (Verb Network Strengthening Treatment; Edmonds, Mammino & Ojeda, 2014), and another (NARNIA; Whitworth et al.,2015) used the treated sentences to build narratives.

In the Communication category there were 13 studies (12%) aimed at improving the participant's ability to convey a message. There were eight examples of CIAT (Constraint Induced Aphasia Therapy) type intervention (for example, Pulvermuller et al., 2001), and five examples of PACE type (Promoting Aphasics' Communicative Effectiveness (Davis, 2005)).

Eight studies (7%) were included in the conversation category. This category referred to any non-specific intervention carried out at conversational level. Four studies described SLT led conversation practice and in the other four cases volunteer visitors conversed with the participants.

The MIT category (N=6) included five studies utilizing Melodic Intonation Therapy (Albert, Sparks & Helm, 1973) and another study describing a similar intonation-based intervention.

Few therapies for reading were included (n=4). There were two oral reading interventions, and two computer-based reading comprehension interventions.

There were two studies categorised as “other”. The study by Freed, Marshall & Nippold (1995) was an associate learning task. It assessed PWA’s ability to learn to associate words to abstract symbols. It contrasted success in naming the symbols depending on whether the participants were given cues or developed personalised cues. The other study, by Nobis-Bosch, Springer, Radermacher & Huber (2011) looks at the use of barcoded pictures to allow the participants to be able to practice repeating dialogues.

The percentage of examples of each intervention category is shown in Figure 2. In summary, the largest category of intervention was Lexical, the majority of which were interventions for word naming. There were only six studies to improve word comprehension. Thirty-two (29%) of the interventions could be classified as general and another eight (7%) were unspecified conversation practice. Only six (5%) studies described therapies to improve sentence processing and four (4%) therapies for improving reading.

Figure 2 about here

Studies categorized by research design

Analyses looked at the types of research design (see method for an explanation of the types of design), the use of standardized tests to measure outcomes, and the numbers of participants; all measures being an indication of the quality and the robustness of the efficacy studies being reviewed. Parallel designs (comparing two interventions) are the most numerous (n= 35, 38%) followed by RCTs (n=24, 26%) and case series (n=21, 23%), with the more complicated cross-over design being the least represented in the sample (n=13, 14%). The majority of RCT and parallel studies included at least one published, standardized test in their outcome measures (81% and 90%

respectively), while this number drops to under 50% for cross-over and case study designs (45% and 46%). The number of studies using each design is shown in Table 2

Table 2 about here

Since we were looking at studies published over a number of decades, it was interesting to see if the different types of research designs had changed in popularity over time. For each research design, the number of publications were counted for each decade from 1980 to the present. The small number of studies from the late 1970s were not included in the analysis. Table 3 plots the number of each research design across the four decades. Across the first three decades there was no difference in the number of RCT studies (4 each decade). There were just one or two cross-over designs for each of these decades. The parallel design studies showed a steady rise while the case studies did not appear until the 2000s. More studies of all designs were published in the 2010s, despite being measured across only part of the decade.

Table 3 about here

There was a large variation in the number of participants in the 93 studies. The mean number was 33 with the median (15) and the standard deviation (47) suggesting a large degree of variance. It should be noted that the figures given are based on the total number of participants recruited in each study, rather than numbers for the different arms of the study. The smallest number of participants was artificially set by our exclusion criteria; any studies with fewer than four participants were rejected. The maximum number of participants was 281. Eighteen studies each had fewer than 10 participants. The majority of these were case series designs; this applied to 12 of the 21 studies with this design. However, there were also three cross-over studies and three parallel studies which described studies

with fewer than 10 participants each. Only 16 studies comprised studies with more than 50 participants, and only seven studies included more than 100 participants.

The mean number of participants for each design differs greatly; the mean no of participants for RCT and parallel designs are 66 and 32 respectively. By contrast the cross-over and case series designs had a mean of 11 and 9 participants. The number of participants differs significantly across the designs (Kruskal Wallis Test: $K(3) = 35.98$, $p<.00001$). However, the large standard deviations denote a degree of overlap.

Participant selection criteria across studies

Since participant selection concerning the aphasic profile is an important parameter for replication and implementation, participant selection criteria were also considered. Studies which specified the selection of participants on the basis of specific (cognitive) linguistic tests, or on scores on subtests of test batteries (not the test batteries as a whole) were categorized as having specific selection criteria. Only 39 studies (42%) did so, with the other 54 studies (58%) not specifying specific deficits. Of the 39 studies, ten studies (11%) specified non-fluent aphasia as an inclusion criterion and clarified this with test criteria, although there was no consensus across the studies in terms of how non-fluency was diagnosed. A further two studies specified non-fluent aphasia as an inclusion criterion but gave no detail about the test scores used to make this diagnosis, so these two studies were in the group of 54 studies categorised as having unspecified selection criteria.

Of the 39 studies where some aphasia selection criteria were given, 37 specified performance on lexical measures (word comprehension, naming, repetition). One MIT study assessed singing ability in addition to a lexical measure. One study specified inclusion on the basis of sentence processing

ability, and one based on a measure of global severity. The majority of studies which specified the selection of participants (26/37) were studies of lexical intervention; and 5/37 were sentence treatments, 3/37 were MIT, 2/37 were communication treatments and there was one instance of each of the following intervention types: conversation, general and other.

Was there any relationship between category of intervention and choice of research design? Because numbers were small (>14 studies) in many of the categories of intervention, it was decided to only include the two largest categories in this analysis. These were the categories General (32 studies) and Lexical (40 studies). The studies using these two approaches were selected and the proportion of design types for each of the therapies were calculated and graphed (Figure 3). Visual inspection of Figure 3 indicates that a clear pattern emerges; general therapies most commonly use a parallel or RCT design, whereas lexical intervention is more associated with case studies.

Figure 3 about here

Quality of reporting across studies

Was there a difference of quality of reporting across the different research designs as measured by the TIDieR checklist? The proportion of studies which provided information for each of the TIDieR checklist items, is given in Figure 4. The majority of studies provided enough information about for eight of the 12 categories to be given a score of one in each of those categories (as described in the method, the scoring was binary with a score of one given if there was any information given for an item on the checklist). Information was most often missing on the location of the intervention and how the intervention might have been individualized or modified through the course of the study. Equally lacking were considerations of measuring fidelity and adherence. This pattern held across the

four types of design; A Kruskal- Wallis test indicated that there were no significant differences in the mean TIDieR scores for each study between the four designs.

Figure 4 about here

Did the novel quality-of-therapy-reporting measure (QTR) give better information on the therapy descriptions? A majority of studies (71/93, 76%) described the general method given and the details of the service delivery used (86/93, 86%). By contrast, only 37 (40%) gave sufficient information about the content in terms of materials and items, 48 about the responses required (52%), and 46 information on dealing with incorrect responses (49%). Just 19 (20%) of the 91 studies achieved the full score of 5. To exemplify the scoring procedure, consider the following two papers. Abel et al. (2014) scored for the materials and items question because they reported their stimulus items and material as follows, " 132 pictures from Snodgrass and Vanderwart (1980) …were used. 90 picture names with the lowest baseline performance were attributed to experimental sets of 30 items each for semantic therapy (SEM), phonological therapy (PHO), and untrained control (CON) … [items] were controlled for comparable performance during baseline as well as linguistic parameters of spoken lemma frequency (CELEX German Database,2001), visual complexity and familiarity of pictures(Genzel, Kerkhoff,&Scheffter,1995… the number of syllables … we also attempted to balance various semantic fields.", page 156. Bakheit et al. (2007) did not score on this question because they reported their stimulus items as follows, "Therapy exercises targeted improvement in understanding and expression both of spoken and written language in order to improve communication in everyday life. These included tasks such as picture/object selection, naming objects, describing and recognizing association between items, facilitating the expression of feelings and opinions and improving conversational skills. Patients were also encouraged to use gesture and other means of non-verbal communication, including a wide range of communication aids and equipment.", page 887.

Table 4 gives the number and proportion of studies deemed to have achieved each of the five indices

Table 4 about here

Two analyses were carried out to look at the relationship between QTR and the research design quality. Is it the case that stronger designs were related to better therapy descriptions? First, we looked to see whether there was any correlation between the QTR measure and the number of participants. using a Pearson rank correlation. A moderate, inverse correlation between QTR and number of participants was found ($r_s = -.514$, $p<.01$). That is, as the quality of therapy reporting increases, the number of participants reduces.

Secondly did the QTR scores differ across the four different research designs? Considering the RCT and parallel designs to be the most robust do these score higher QTRs? The mean scores and standard deviations are shown in Table 5. Similarly to the above inverse finding, the mean QTR score for the RCT designs was 2.13/5 and for the parallel designs it was 2.74. The QTR mean score for cross over designs was somewhat higher at 3.62 and highest of all (4.00) for the case series studies. A Kruskal-Wallis Test was carried out to establish if the QTR differed significantly across the four different designs and was found to be significant ($k(3) = 19.61$ $p= <.001$). Post-hoc t-tests (homogeneity of scores not assumed) indicated significant differences in scores between the RCT designs and the Case studies (two sample t-test: $t(42) =4.53$, $P<.001$ two tailed), as well as between the parallel studies and case studies (two sample t-test: $t(54) =3.27$, $P=.002$ two tailed).

Table 5 about here

Discussion

The purpose of this umbrella review ((Becker, Deeks, Oxman, 2006) was to characterize the types of intervention reported in the aphasia evidence base; to evaluate the quality of the studies (research design, measurement of outcomes, sample size), explore the specification of inclusionary criteria, evaluate the quality of aphasia intervention reporting, and to assess aphasia intervention replicability. Our findings sit in the wider context of concern about reporting quality in aphasia intervention research and recent initiatives to improve this (Brady et al., 2016; Brady et al. 2020).

Acknowledgement of inadequate scientific reporting has resulted in the development of reporting guidelines (Simera, Moher, Hoey, Schulz & Altman, 2010), and other materials aimed at improving scientific reporting (such as the TIDieR checklist). Although there is encouraging evidence that guidelines can improve the quality of reporting (e.g. Cobo et al., 2011) these positive findings arise mainly from RCTs examining relatively straightforward interventions whereas complex behavioural interventions, such as speech and language therapy interventions for aphasia, continue to be less adequately described. This may be because of the complexity of these interventions, or because reporting guidelines tend to apply to a wide range of studies and the meaning and application of terms may not apply equally well across intervention types. For example, the Consolidated Standards of Reporting Trials (CONSORT) for randomized controlled trials recommends that the description of interventions (item 5) should include 'sufficient details to allow replication, including how and when they were actually administered' (Schulz, Altman, & Moher, 2010; p.699). However, this suggestion is unspecific with regard to what details are needed for replication with fidelity. The TIDieR further specifies some generic details needed for replication (e.g. materials, procedures, providers, modes of delivery, location, etc.) but the meaning of these terms varies across intervention types and may be interpreted differently by researchers. A further possibility is that intervention studies are not

adequately described in published studies because of publishing space requirements. A solution to this issue would be to include full details as supplemental material.

In the following sections, our findings relating to intervention type, quality of study design, and selection criteria are first discussed, before the quality of the intervention reporting and replicability is evaluated.

What types of intervention are included in the studies?

The two largest categories of intervention type were Lexical (36%) and General (29%), with other types collectively making-up only 35% of the evidence base. Although the interventions included in the Lexical category include a range of different ways to treat word production, on the whole this review reveals a lack of diversity of intervention approaches in the aphasia evidence base. There was a small number of studies aimed at improving conversation either with specific targets (CILT and PACE) or general conversation in groups, often led by volunteers. There was an even smaller number of studies describing interventions for sentences, intonation and reading. This finding does not reflect the diversity, or the functional specificity, of intervention carried out in clinics. Future studies should therefore encompass a broader range of interventions as well as taking account of the research priorities of people with aphasia, carers and SLT/Ps (Franklin et al., 2018).

There is a lack of specificity in the descriptions of intervention in a large part of this evidence base, hampering interpretation and replicability. It was often unclear what was meant by the non-specific terms used to describe intervention by the studies in the General category (‘usual care’, ‘traditional therapy’, ‘stimulation therapy’, ‘model-based therapy’). Since the content of aphasia intervention is hugely variable (Brady et al., 2016; Brady et al., in press . 2020) this is a significant gap in the evidence base. The combination of interventions and order of different interventions is unexplored

in the evidence base although is commonly used in clinical practice. For example, functional communication intervention may be combined with work on a specific aspect of linguistic impairment or impairment-based intervention may be followed by intervention directed to functional communication. The current evidence base does not provide information about the effects of this combination nor about how to order each component for best effect.

What is the quality of the studies in terms of types of research design and the use of standardized outcome measures?

In this evidence base, parallel group designs are the most numerous, with the more complicated cross-over design being the least represented. The majority of RCT (81%) and parallel studies (90%) included at least one standardized outcome measure. General therapies most commonly use a parallel or RCT design, whereas naming is more associated with multiple case studies. The proportion of case study designs using a standardized outcome measure was less than half (45%). There have been more examples of all design types (except cross-over) published over time. This is particularly true of case series designs, and parallel designs. The increase in the latter is perhaps because of the ethical difficulty of no-treatment controls. However, increased use of parallel designs could be considered problematic where an RCT, or other appropriate design, has not previously shown that at least one of the designs is effective. This is because the design limits the possible interpretation of the results: if both treatments show improvement, the design does not allow the conclusion that either treatment is better than no intervention. Research design is also linked to the phase of the research and the research question. Early phase studies are often case-series, single-subject, and small group design to explore feasibility and early efficacy (i.e., proof of concept). These early study designs are essential for pilot work that lays the foundation for larger, more controlled studies.

What is the quality of studies in terms of sample size, and are aphasia profiles specified in inclusionary criteria?

There was an extremely large variation in the number of participants in the 93 studies, with only 16 studies including more than 50 participants, and only seven studies including more than 100 participants. The conclusion from this is that almost all the group studies designed to look at effectiveness in the aphasia intervention evidence base are likely to have been under-powered. Small numbers effect the degree of experimental control for measuring intervention effects, making it more likely that results are influenced by the effects of observation or spontaneous recovery. In addition, larger participant numbers are needed for post-hoc analysis of candidacy. It is not always clear from the studies reviewed why numbers are small, but it can be hypothesized that availability of patients and funding are contributory factors, as well as the stage of development of the intervention, with early phases (e.g. Phase I, proof of concept) inevitably including small numbers.

We also evaluated the extent to which the studies in this review recruited participants with specific aphasia profiles. Aphasia is a heterogeneous condition, with potentially quite different interventions required for different aphasic difficulties, so it might be expected that there would be very specific descriptions of aphasia profiles for testing specific interventions. However, only 39 studies (42%) included aphasia selection criteria (mostly performance on word comprehension, naming and repetition tests), most commonly those describing lexical interventions. Often information was given post-selection about the classical aphasia type and general severity, but even at this level there were no post-hoc analyses to establish whether people with different aphasia types or severity responded differently to the intervention. Inclusion criteria tended instead to be focused on factors such as age, first stroke or time post-onset, but for these studies the participants with aphasia were treated as an homogenous group in terms of their communication profiles. In terms of acute, post-acute and chronic stages of aphasia, the majority of studies involved patients more than three months post

onset. Only 22% of studies included patients less than three months post onset, despite the fact that there is evidence for the efficacy of treatment for participants with aphasia in all stages of recovery, and indications that outcomes may be greatest in the acute stage (Robey, 1998).

This gap in the evidence base is a cause for concern because aphasia intervention is best described in terms of specific interventions for specific aphasia profiles. For example, there is evidence that efficacy is influenced by the effective targeting of intervention at a specific level of language impairment (e.g. phonological output - Jacquemot, Dupoux, Robotham, & Bachoud-Lévi, 2012; and grammatical morphology - de Aguiar, Bastiaanse, & Miceli, 2016).

What is the quality of intervention reporting, and how replicable are the interventions?

Two analyses (TIDieR and QTR) revealed areas of strength and weakness in the evidence base. More than 50% of studies provided information about eight of the twelve TIDieR categories (name, why, materials, procedures, who, how, when, and tailoring), although it should be noted that the rating did not require complete information. Information was most often missing on the location of the intervention, perhaps because this is seen as less important for replicability and implementation. Details about how the intervention might have been modified through the course of the study (modification and actual fidelity) was lacking. Although information about personalisation and adaptation (tailoring) was present in the majority of studies, it was missing in 43/93 studies (46%). Lack of tailoring (and reporting of this) may arise specifically because researchers are aiming for replicability. On the novel quality-of-therapy-reporting measure (QTR), a majority of studies described the general method and the details of the service delivery used but only around a half of studies provided sufficient information for the other indices (intervention content, responses required and type of feedback given). Just 19 (20%) of the 93 studies achieved the full score of five points which would indicate they were fully replicable. There are some studies not achieving a

perfect replicability score where the intervention detail is published more fully elsewhere (such as in a protocol paper on additional online resources elsewhere), however these studies were in the minority. And even where the intervention method is adequately described within the restrictions of the guidelines of the journal, in most cases, the detailed information necessary to perform a reliable replication study was missing.

There was a relationship between the reporting quality rating (QTR), the research design, and the number of participants, indicating that case series studies (those with fewer participants) scored significantly higher on the quality of intervention reporting indices than RCT and parallel designs (with more participants). This reflects the advantages and limitations of different designs outlined by Best and colleagues (Best, Wei Sze, Edmundson, & Nickels, 2019).

Limitations

The studies included in this review were not sourced directly from databases but from the reference lists of other reviews, which may limit confidence in the representativeness of the evidence reviewed here. However, each reference list source, represents reliable, comprehensive and up-to-date searches and collectively they constitute most reputable syntheses available in aphasia rehabilitation research. In addition, our research questions allowed us to focus on the core evidence base in aphasia rehabilitation because we selected for a range of robust study deigns (rather than just RCTs) and for only those studies investigating speech and language therapy interventions in aphasia.

The scope of this review was an analysis of aphasia intervention studies and did not include intervention descriptions published separately (online, in clinical tools, or as a separate trial protocols). This means that our evaluation of how replicable and implementable this evidence base is limited to intervention studies only.

Conclusions

The results of this review revealed reporting gaps in the evidence base that should be targeted as a priority in future research to promote replicability. We would suggest that the field needs both work using case series and work using the larger research designs (Best, et al., 2019). Future research using case series should focus on maximizing their robustness and intervention reporting quality. Systematic reviews and meta-syntheses of good quality single case studies and case series would also strengthen the field and increase our knowledge concerning the effectiveness of specific interventions for specific aphasias (e.g. Pierce, Menahemi-Falkov, O'Halloran, Togher, & Rose 2017). Large, well-powered RCTs with good assessment of participants, replicable description of intervention should crucially include post-hoc regression to look at efficacy of different interventions for different aphasia profiles. Agreed outcome measures, such as the Core Outcome Set proposed by Wallace and colleagues (Wallace et al.,2019, (with recognition of the possibilities in different countries and languages), would then allow for greater comparability, and for post-hoc regression analyses to explore efficacy and candidacy across as well as within studies.

REFERENCES

1. Abel, S., Weiller C., Huber, W. & Willmes K. (2014). Neural underpinnings for model-oriented therapy of aphasic word production. *Neuropsychologia, 57*, 154–165.
2. Adrian, J. A., Gonzalez, M., Buiza, J. J. & Sage, K. (2011). Extending the use of Spanish Computer-assisted Anomia Rehabiltation Program (CARP-2) in people with aphasia. *Journal of Communication Disorders, 44*, 666-677.
3. Aftonomos, L. B., Appelbaum, J. S., & Steele, R. D. (1999). Improving outcomes for persons with aphasia in advanced community-based treatment programs. *Stroke, 30 (7)*, 1370-1379.
4. Agostini, M., Garzon, M., Benavides-Varela, S., De Pellegrin, S., Benicini G., Rossi, G…….Tonin, P. (2014). Telerehabilitation in poststroke anomia. *Biomed Research International,* article ID 706909
5. 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.
6. Albert, R., Sparks, R. & Helm, N. (1973). Melodic Intonation Therapy for Aphasia, *Archives of Neurology, 29(2),* 130-1
7. Albyn Davis, G. (2005). PACE revisited. *Aphasiology, 19(1),* 21-38.
8. Altmann, L. J. P., Hazamy, A. A., Carvajal, P. J., Benjamin, M., Rosenbek, J.C. & Crosson B. (2014). Delayed stimulus-specific improvements in discourse following anomia treatment using an intentional gesture. *Journal of Speech and Hearing Research, 57*, 439-454.
9. Bakheit, A.M.O., Shaw, S., Barrett, L., Wood, J., Carrington, S., Griffiths, S., …..Koutsi, F.. (2007). A prospective, randomized, parallel group, controlled study of the effect of intensity of speech and language therapy on early recovery from poststroke aphasia. *Clinical Rehabilitation, 21*, 885–894

10. Barthel, G., Meinzer, M., Djundja, D., & Rockstroh, B. (2008). Intensive language therapy in chronic aphasia: Which aspects contribute most?. *Aphasiology, 22 (4)*, 408-421.
11. Basso, A., Capitani, E., & Vignolo, L. A. (1979). Influence of rehabilitation on language skills in aphasic patients: A controlled study. *Archives of Neurology*, *36(4)* , 190-196.
12. Becker, L., Deeks, J., & Oxman, A. (2006). Introduction to umbrella reviews. In: *Come to the craic. Abstracts of the 14th Cochrane Colloquium* 23-26 Oct, Dublin, UK.
13. Benjamin, M.L., Towler, S., Garcia, A., Park, H., Sudhyadhom, A., Harnish, S. …..Crosson, B. (2014). A behavioral manipulation engages right frontal cortex during aphasia therapy. *Neurorehabilitation and Neural Repair*, *28,* 545-553
14. Best, W., Greenwood, A., Grassly, J. & Hickin, J. 2008. Bridging the gap:can impairment-based therapy for anomia have an impact at the psycho-social level *International Journal of Language and Communication Disorders, 43(4)*, 390-407.
15. 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.
16. Best, W., Sze, W. P., Edmundson, A. & Nickels, L. (2019). What counts as evidence? Swimming against the tide: Valuing both clinically informed experimentally controlled case series and ramdomized controlled trials in intervention research. *Evidence-Based Communication Assessment and Intervention, 13(3)*, 107-135.
17. Bonifazi, S., Tomaiuolo, F., Altoè, G., Ceravolo, M. G., Provinciali, L., & Marangolo, P. (2013). Action observation as a useful approach for enhancing recovery of verb production: new evidence from aphasia. *European Journal of Physical and Rehabilitation Medicine, 49,* 473-481.
18. Bowen, A., Hesketh, A., Patchick, E., Young, A., Davies, L., Vail, A…..Tyrell, 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.

19. Bowen, A., Hesketh, A., Patchick, E., Young, A., Davies, L., Vail, A.,….Tyrell, P. (2012). Effectiveness of
enhanced communication therapy in the first four months after stroke for aphasia and dysarthria: a randomised controlled trial, BMJ, ,345, e4407
20. Brady, M.C., Kelly, H., Godwin, J., Enderby, P. & Campbell P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews* , 1(6), Article No.: CD000425.
21. Brady, M., Ali, M., VandenBerg, K., Williams, L.J., Williams, L.R., Abo Harris Wright, H. (2020 – in press). Communicating simply, but not too simply – reporting of participants and speech and language interventions for aphasia after stroke. *International Journal of Speech Language Pathology*.
22. Brindley, P., Copeland, M., Demain, C., & Martyn, P. (1989). A comparison of the speech of ten chronic Broca's aphasics following intensive and non-intensive periods of therapy. *Aphasiology, 3*(8), 695-707.
23. Byng, S., Nickels, L. & Black, M. (1994). Replicating therapy for mapping deficits in agrammatism: Remapping the deficit? *Aphasiology, 8(4),* 315-341.
24. Campbell, M., Katikireddi, S. V., Hoffmann, T., Armstrong, R., Waters, E., & Craig, P. (2018). TIDieR-PHP: a reporting guideline for population health and policy interventions. *British Medical Journal* , *361*, k1079.
25. Carragher, M., Sage, K. & Conroy, P. (2015). Outcomes of treatment targeting syntax production in people with Broca's type aphasia: evidence from psycholinguistic assessment tasks and everyday conversation. *International Journal of Language and Communication Disordes, 50,* 322-336.

26. Carragher, M., Sage, K., & Conroy, P. (2013). The effects of verb retrieval therapy for people with non-fluent aphasia: Evidence from assessment tasks and conversation. *Neuropsychological Rehabilitation, 23(6),* 846-887.
27. Caute, A., Pring, T., Cocks, N., Cruice, M., Best, W., & Marshall, J. (2013). Enhancing communication through gesture and naming therapy. *Journal of Speech, Language, and Hearing Research, 56,* 337-31
28. Chan, A.W., Tetzlaff, J.M., Altman, D.G., Laupacis, A., Gøtzsche, P.C., Krleža-Jerić, K., ….Doré, C.J. (2013). SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Annals of Internal Medicine*, *158*(3), 200-207.
29. Cherney, L. R. (2010). Oral reading for language in aphasia (ORLA): Evaluating the efficacy of computer-delivered therapy in chronic nonfluent aphasia. *Topics in Stroke Rehabilitation, 17(6),* 423-431.
30. Ciccone, N., West, D., Cream, A., Cartwright, J., Rai, T., Granger, A., …… Godecke, E. (2016). Constraint-induced aphasia therapy (CIAT): a randomised controlled trial in very early stroke rehabilitation. *Aphasiology, 30(5*), 566-584.
31. Cobo, E., Cortes, J., Ribera, J. M., Cardellach, F., Selva-O'Callaghan, A., Kostov, B. …Vilardell, K. (2011). Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomised trial. *British Medical Journal*,343,1–11.
32. Code, C., Torney, A., Gildea-Howardine, E. & Willmes, K. (2010).Outcome of a one-monthy therapy intensive dose chronic aphasia: variable individual responses, *Seminars in Speech and Language*, 31 (1) 21-33.
33. Conklyn, D., Novak, E., Boissy, A., Bethoux, F. & Chemali, K. (2012). The effects of modified melodic intonation therapy on nonfluent aphasia: a pilot study. *Journal of Speech, Language and Hearing Research, 55,* 1463-1471.

34. Conroy, P., Sage, K. &Lambon Ralph, M. A. (2009). Errorless and erroful therapy for verb and noun naming in aphasia. *Aphasiology, 23,* 1311-1337.
35. Cotterill, S., Knowles, S., Martindale, A. M., Elvey, R., Howard, S., Coupe, N., Wilson, P., & Spence, M. (2018). Getting messier with TIDieR: embracing context and complexity in intervention reporting. *BMC medical research methodology, 18(1),* 12.
36. Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2011). Developing and evaluating complex interventions. Medical Research Council, UK.
37. Crerar, M. A., Ellis, A. W., & Dean, E. C. (1996). Remediation of sentence processing deficits in aphasia using a computer-based microworld. *Brain and Language, 52(1)*, 229-275.
38. David, R., Enderby, P., & Bainton, D. (1982). Treatment of acquired aphasia: speech therapists and volunteers compared. *Journal of Neurology, Neurosurgery & Psychiatry, 45(11),* 957-961.
39. De Jong-Hagelstein, M., Van de Sandt-Koenderman, M., Prins, N., Dippel, D., Koudstaal, P.J. & Visch-Brink, E.G. (2011). Efficacy of early cognitive-linguistic treatment and communicative treatment in aphasia after stroke: a randomised controlled trial (RATS-2). *Journal of Neurology Neurosurgery and Psychiatry, 82*, 399-404.
40. Denes, G., Perazzolo, C., Piani, A., & Piccione, F. (1996). Intensive versus regular speech therapy in global aphasia: A controlled study. *Aphasiology, 10(4),* 385-394.
41. Des Roches, C. A., Balachandran, I., Ascenso, E.M., Tripodis, Y. & Kiran, S. (2015). Effectiveness of an impairment-based individualized rehabilitation program using an iPad-based software platform. *Frontiers in Human Neuroscience, 8,* Article 1015.
42. Doesborgh, S., van de Sandt-Koenderman, M., Dippel, D., van Harskamp, F., Koudstaal, P., & 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,*141-146.

43. Doesborgh, S., van de Sandt-Koenderman, M., Dippel, D., van Harskamp, F., Koudstaal, P., & Visch-Brink, E.G. (2004). Cues on request: The efficacy of Multicue, a computer program for wordfinding therapy. *Aphasiology, 18(3)*, 213-222.

44. Drummond, S. S., & Rentschler, G. J. (1981). The efficacy of gestural cueing in dysphasic word-retrieval responses. *Journal of Communication Disorders, 14(4),* 287–98.

45. Edmonds, L. A., Mammino, K., & Ojeda, J. (2014). Effect of verb network strengthening treatment (VNeST) in persons with aphasia: Extension and replication of previous findings. *American Journal of Speech-Language Pathology, 23(2),* 312-329.

46. Elman, R. J., & Bernstein-Ellis, E. (1999). The efficacy of group communication treatment in adults with chronic aphasia. *Journal of Speech, Language, and Hearing Research, 42(2)*, 411-419.

47. Faltynek, P., Salter, K., Cotoi, A., Macaluso, S., Orange, J. & Teasell, R. (2018) Evidence-Based Review of Stroke Rehabilitation: Aphasia and Apraxia (18th edition) *downloaded from http://www.ebrsr.com/ March 2018.*

48. Ferguson, N.F., Evans, K. & Raymer, A.M. (2012). A comparison of intention and pantomime gesture treatment for noun retrieval in people with aphasia*. American Journal of Speech-Language Pathology, 21,* S126-S139.

49. Fillingham, J., Sage, K. & Lambon Ralph, M. (2005). Further explorations and an overview of errorless and errorful therapy for aphasic word-finding difficulties: The number of naming attempts during therapy affects outcome. *Aphasiology, 19,*597-614.

50. Franklin, S., Harhen, D., Hayes, M., Demos Mc Manus, S. & Pollock, A. Top 10 research priorities relating to aphasia following stroke. *Aphasiology, 32(11),* 1388-1395.

51. Freed, D. B., Marshall, R. C., & Nippold, M. A. (1995). Comparison of personalized cueing and provided cueing on the facilitation of verbal labelling by aphasic subjects. *Journal of Speech, Language, and Hearing Research, 38(5),* 1081-1090.

52. Fridriksson, J., Baker, J. M., Whiteside, J., Eoute Jr, D., Moser, D., Vesselinov, R., & Rorden, C. (2009). Treating visual speech perception to improve speech production in nonfluent aphasia. *Stroke, 40(3),* 853-858.
53. Godecke, E., Hird, K., Lalor, E. E., Rai, T., & Phillips, M. R. (2012). Very early poststroke aphasia therapy: a pilot randomized controlled efficacy trial. International Journal of *Stroke, 7(8),* 635-644.
54. Harnish, S.M., Morgan, J., Lundine, J.P., Bauer, A., Singletary, F., Benjamin, M.L., Gonzalez Rothi, L.J. & Crosson, B. (2014). Dosing of a cued picture-naming treatment for anomia. *American Journal of Speech-Language Pathology, 23(2)*, S285-S299.
55. Hickin, J., Best, W., Herbert, R., Howard, D. & Osborne, F. (2002). Phonological therapy for word-finding difficulties: A re-evaluation. *Aphasiology, 16*, 981-999
56. Hinckley, J. J., & Craig, H. K. (1998). Influence of rate of treatment on the naming abilities of adults with chronic aphasia. *Aphasiology, 12(11),* 989-1006.
57. Hinckley, J. J., Patterson, J. P., & Carr, T. H. (2001). Differential effects of context-and skill-based treatment approaches: Preliminary findings. *Aphasiology, 15(5)*, 463-476.
58. Hinckley, J., & Carr, T. (2005). Comparing the outcomes of intensive and non-intensive context-based aphasia treatment. *Aphasiology, 19(10-11),* 965-974.
59. Hoffmann, T. C., Glasziou, P. P., Bouton, I., Milne, R., Perera, R., Moher, D…..Michie, S. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR checklist and guide. *British Medical Journal, 348*, g 1687.
60. Howard, D., Patterson, K., Franklin, S., Orchard-Lisle, V., & Morton, J. (1985). Treatment of word retrieval deficits in aphasia. *Brain, 108(8)*, 17-829.
61. Jacquemot, C, Dupoux, E., Robotham, L. & Bachoud-Lévi (2012). Specificity in rehabilitation of word production: a meta-analysis and a case study. *Behavioral Neurology, 25(2),* 73-101.

62. Katz, R. C. & Wertz, R. T.(1992). Computerized hierarchical reading treatment in aphasia. *Aphasiology, 6*,165-177

63. Katz, R. C., & Wertz, R. T. (1997). The efficacy of computer-provided reading treatment for chronic aphasic adults. *Journal of Speech, Language, and Hearing Research, 40(3)*, 493-507.

64. Kiran, S, Meier, E. L., Kapse, K.J. & Glynn, P.A. Changes in task-based effective connectivity in language networks following rehabilitation in post-stroke patients with aphasia. *Frontiers in Human Neuroscience, 9*, Article 316.

65. Kurland, J., Stanek, E. J., III, Stokes, P., Li, M., & Andrianopoulos, M. (2016). Intensive language action therapy in chronic aphasia: A randomized clinical trial examining guidance by constraint. *American Journal of Speech-Language Pathology, 25(4S)*, S798-S812.

66. Kurland, J., Wilkins, A. R., & Stokes, P. (2014). iPractice: Piloting the effectiveness of a tablet-based home practice program in aphasia treatment. In: *Seminars in Speech and Language, 35 (1),* 51-64. NIH Public Access.

67. Lacey, E. H., Lott, S. N., Snider, S. F., Sperling, A., & Friedman, R. B. (2010). Multiple oral re-reading treatment for alexia: the parts may be greater than the whole. *Neuropsychological Rehabilitation, 20(4)*, 601-623.

68. Laska, A. C., Kahan, T., Hellblom, A., Murray, V., & von Arbin, M. (2011) A randomized controlled trial on very early speech and language therapy in acute stroke patients with aphasia. *Cerebrovascular Diseases*, *1*,66-74.

69. Leonard, C., Laird, L., Burianová, H., Graham, S., Grady, C., Simic, T. & Rochon, E. (2015). Behavioural and neural changes after a "choice" therapy for naming deficits in aphasia: preliminary findings. *Aphasiology, 29(4),* 506-525.

70. Lim, K. B., Kim, Y. K., Lee, H. J., Yoo, J., Hwang, J. Y., Kim, J. A. & Kim, S. K. (2013). The therapeutic effect of neurologic music therapy and speech language therapy in post-stroke aphasic patients. *Annals of Rehabilitation Medicine, 37(4),* 556-562.

71. Lincoln, N. B., Pickersgill, M. J., Hankey, A. I. & Hilton, C. R. (1982). An evaluation of operant training and speech therapy in the language rehabilitation of moderate aphasics. *Behavioural Psychotherapy, 10*, 162-178

72. Lincoln, N. B., & Pickersgill, M. J. (1984). The effectiveness of programmed instruction with operant training in the language rehabilitation of severely aphasic patients. *Behavioural and Cognitive Psychotherapy, 12(3)*, 237-248.

73. Lincoln, N. B., Mulley, G. P., Jones, A. C., McGuirk, E., Lendrem, W., & Mitchell, J. R. A. (1984). Effectiveness of speech therapy for aphasic stroke patients: a randomised controlled trial. *The Lancet, 323(8388),* 1197-1200.

74. Lyon, J.G., Cariski, D., Keisler, L., Rosenbek, J., Levine, R., Kumpula, J., ….. Blanc, M., (1997). Communication partners: Enhancing participation in life and communication for adults with aphasia in natural settings. *Aphasiology, 11(7),* 693-708.

75. Marshall, J., Best, W., Cocks, N., Cruice, M., Pring, T. , Bulcock, G., …. Caute, A. (2012). Gesture and naming therapy for people with severe aphasia: a group study. *Journal of Speech, Language and Hearing Research, 55,*726-738.

76. Martins, I. P., Leal, G., Fonseca, I., Farrajota, L., Aguiar, M., Fonseca, J. ……Ferro, J. M., et al. (2013). A randomized, rater-blinded, parallel trial of intensive speech therapy in sub-acute post-stroke aphasia: the SP-I-R-IT study. *International Journal of Language and Communication Disorders,48(4),* 421-31.

77. Mattioli, F., Ambrosi, C., Mascaro, L., Scarpazza, C., Pasquali, P., Frugoni, M.,……Gasparotti, R. (2014). Early aphasia rehabilitation is associated with functional reactivation of the left inferior frontal gyrus: a pilot study. *Stroke, 45(2),* 545-552.

78. Meikle, M., Wechsler, E., Tupper, A., Benenson, M., Butler, J., Mulhall, D., & Stern, G. (1979). Comparative trial of volunteer and professional treatments of dysphasia after stroke. *British Medical Journal, 2,* 87-89.

79. Meinzer, M., Streiftau, S. & Rockstroh, B. (2007). Intensive language training in the rehabilitation of chronic aphasia - effective training by laypersons. *Journal of the International Neuropsychological Society,13(5),*846-53.

80. Meinzer, M., Djundja, D., Barthel, G., Elbert, T. & Rockstroh, B. (2005). Long-term stability of improved language functions in chronic aphasia after constraint-induced aphasia therapy. *Stroke, 36(7),* 1462-1466.

81. Menke, R., Meinzer, M., Kugel, H., Deppe, M., Baumgaertner, A., Schiffbauer, H. …, Breitenstein, C. (2009). Imaging short- and long-term training success in chronic aphasia. *BMC Neuroscience 10*, 118.

82. Nobis-Bosch, R., Springer, L., Radermacher, I., & Huber, W. (2011). Supervised home training of dialogue skills in chronic aphasia: A randomized parallel group study. *Journal of Speech, Language, and Hearing Research, 54,* 1118-1136..

83. Palmer, R., Enderby, P., Cooper, C., Latimer, N., Julious, S., Paterson, G.… Delaney, A. (2012). Computer therapy compared with usual care for people with long-standing aphasia poststroke: a pilot randomized controlled trial. *Stroke, 43(7)*,1904-1911.

84. Pierce, J. E., Menahemi-Falkov, M., O'Halloran, R., Togher, L. & Rose, M. L. (2019) Constraint and multimodal approaches to therapy for chronic aphasia: A systematic review and meta-analysis, *Neuropsychological Rehabilitation, 29(7)*, 1005-1041.

85. Prins, R. S., Schoonen, R. & Vermeulen J. (1989.) Efficacy of two different types of speech therapy for aphasic stroke patients. *Applied Psycholinguistics, 10*,85-123.

86. Pulvermüller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., & Taub, E. (2001). Constraint-induced therapy of chronic aphasia after stroke. *Stroke, 32(7),* 1621-1626.

87. Raglio, A., Oasi, O., Gianotti, M., Rossi, A., Goulene, K., & Stramba-Badiale, M. (2016). Improvement of spontaneous language in stroke patients with chronic aphasia treated

with music therapy: A randomized controlled trial. *International Journal of Neuroscience, 126(3),* 235-242

88. 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.
89. Rochon, E., Laird, L., Bose, A., & Scofield, J. (2005). Mapping therapy for sentence production impairments in nonfluent aphasia. *Neuropsychological Rehabilitation, 15(1),* 1-36.
90. Rose, M. L., Attard, M. C., Mok, Z., Lanyon, L. E., & Foster, A. M. (2013). Multi-modality aphasia therapy is as efficacious as a constraint-induced aphasia therapy for chronic aphasia: A phase 1 study. *Aphasiology*, *27(8)*, 938-971.
91. Sandberg, C. & Kiran, S. (2014). How justice can affect jury: Training abstract words promotes generalisation to concrete words in patients with aphasia. *Neuropsychological Rehabilitation, 24(5),* 738-769.
92. Schulz, K. F., Altman, D. G., & Moher, D.; CONSORT Group.(2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *British Medical Journal, 340*, c332.
93. Seron, X., Deloche, G., Bastard, V., Chassin, G., & Hermand, N. (1979). Word-finding difficulties and learning transfer in aphasic patients. *Cortex, 15(1),* 149-155.
94. Shewan, C. M. & Kertesz, A. (1984). Effects of Speech and Language Treatment on Recovery from Aphasia. *Brain and language, 23,*272-299.
95. Sickert, A., Anders, L. C., Münte, T. F., & Sailer, M. (2014). Constraint-induced aphasia therapy following sub-acute stroke: a single-blind, randomised clinical trial of a modified therapy schedule. *Journal of Neurology, Neurosurgery and Psychiatry, 85(1),* 51-55.
96. Simera, I., Moher, D., Hoey, J., Schulz, K. F., & Altman, D. G. (2010). A catalogue of reporting guidelines for health research. *European Journal of Clinical Investigation, 40(1),* 35-53.

97. Smania, N., Aglioti, S. M., Girardi, F., Tinazzi, M., Fiaschi, A., Cosentino, A., & Corato, E. (2006). Rehabilitation of limb apraxia improves daily life activities in patients with stroke. *Neurology, 67(11),* 2050-2052.

98. Snell, C., Sage, K.,& Lambon Ralph, M. A. (2010). How many words should we provide in anomia therapy? A meta-analysis and a case series study. *Aphasiology 24(9)*1064-1094.

99. Stahl, B., Henseler, I., Turner, R., Geyer, S. & Kotz, S. A. (2013). How to engage the right brain hemisphere in aphasics without even singing: evidence for two paths of speech recovery. *Frontiers in Human Neuroscience, 7*, Article 35.

100. Stahl, B., Mohr, B., Dreyer, F. R., Lucchese, G., & Pulvermüller, F. (2016). Using language for social interaction: Communication mechanisms promote recovery from chronic non-fluent aphasia. *Cortex*, *85,* 90-99.

101. Takizawa, T., Nishida, N., Ikemoto, A. & Kurauchi, N. (2015). Comparison of single-word therapy versus sentence therapy for verb retrieval and sentence production in Broca's aphasia. *Aphasiology, 29(2),* 169-194.

102. Van der Meulen, I., Van De Sandt-Koenderman, M. W., Heijenbrok, M. H., Visch-Brink, E., & Ribbers, G. M. (2016). Melodic Intonation Therapy in Chronic Aphasia: Evidence from a Pilot Randomized Controlled Trial. *Frontiers in Human Neuroscience*, *10,* 533.

103. Van der Meulen, I., van de Sandt-Koenderman, W. M. E., Heijenbrok-Kal, M. H., Visch-Brink, E. G., & Ribbers, G. M. (2014). The efficacy and timing of melodic intonation therapy in subacute aphasia. *Neurorehabilitation and Neural Repair, 28(6)*, 536-544.

104. van Hees, S., Angwin, A., McMahon K. & Copland, D. (2013). A comparison of semantic feature analysis and phonological components analysis for the treatment of naming impairments in aphasia. *Neuropsychological Rehabilitation, 23(1),* 102-132.

105. Wallace, S.J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K. …. Copland, D. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International journal of stroke*, *14(2),* 180-185

106. Wan, C. Y., Zheng, X., Marchina, S., Norton, A. & Schlaug, G. (2014). Intensive therapy induces contralateral white matter changes in chronic stroke patients with Broca's aphasia. *Brain and Language, 136,*1-7.

107. Warren, S.F., Fey, M. E. & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disability Research Reviews, 13(1),* 70-77.

108. Wertz, R.T., Weiss, D.G., Aten, J.L., Brookshire, R.H., Garcia-Bunuel, L., Holland, A.L.….Greenbaum, H., (1986). Comparison of clinic, home, and deferred language treatment for aphasia: A Veterans Administration cooperative study. *Archives of Neurology, 43(7),* pp.653-658

109. Whitworth, A., Leitao, S., Cartwright, J., Webster, J., Hankey, J. & Zach, D. (2015). NARNIA: a new twist to an old tale. A pilot RCT to evaluate a multilevel approach to improving discourse in aphasia. *Aphasiology, 29,*1345-1382.

110. Whyte, J. (2008). A grand unified theory of rehabilitation (we wish!). The 57th John Stanley Coulter Memorial Lecture. *Archives of Physical Medicine and Rehabilitation*, *89*(2), 203-209.

111. Wilssens, I., Vandenborre, D., van Dun, K., Verhoeven, J., Visch-Brink, E., & Mariën, P. (2015). Constraint-induced aphasia therapy versus intensive semantic treatment in fluent aphasia. *American Journal of Speech-Language Pathology, 24(2),* 281-294.

112. Woolf, C., Caute, A., Haigh, Z., Galliers, J., Wilson, S., Kessie, A., Hirani, S., Hegarty, B. & Marshall, J. (2016). A comparison of remote therapy, face to face therapy and an attention control intervention for people with aphasia: a quasi-randomised controlled feasibility study. *Clinical Rehabilitation, 30(4),* 359-373.

113. Woolf, C., Panton, A., Rosen, S., Best, W. & Marshall, J. (2014). Therapy for auditory processing impairment in aphasia: An evaluation of two approaches. *Aphasiology, 28(12)*, 1481-1505.

114. Worrall, L., & Yiu, E. (2000). Effectiveness of functional communication therapy by volunteers for people with aphasia following stroke. *Aphasiology, 14(9)*, 911-924.

115. Yamoto, T., Maher, C., Saragiotto, B., Moseley, A., Hoffmann, T., Elkins, M. & Jette, A. (2016). The TIDieR checklist will benefit the physical therapy profession. *Physical Therapy, 96(7)*, 930-1

Table 1 Items in the TIDieR[1] (Hoffmann et al., 2014)

Brief name	When and how much Number of sessions, duration and intensity
Why Rationale or goal of the intervention	Tailoring Any adaptations for the individual?
What – Materials Materials used in intervention	Modifications Changes made to the intervention during the course of the study
What – Procedures Each procedure with enabling activities	How well (fidelity and adherence) – Planned How were fidelity or adherence assessed?
How (mode of delivery) Face to face or other, individual or group	How well (fidelity and adherence) – Actual Extent to which intervention was delivered as planned
Where Location where intervention occurred	

[1] The Template for Intervention Description and Replication (TIDieR) (Hoffmann et al. 2014)

Table 2 Number of studies per design (n=93 studies)

	No of studies	(%)	% used at least 1 published test in outcome measures
Parallel	**35**	38	81
Randomised Controlled Trial	**24**	26	90
Case series	**21**	23	45
Cross over	**13**	14	46

Table 3: No of studies by design over time (n=93 studies)

	1980s	*1990s*	*2000s*	*2010s*
Randomised Controlled Trial	*4*	*4*	*4*	*11*
Parallel	*4*	*6*	*8*	*16*
Cross-over	*2*	*1*	*2*	*8*
Case study	*0*	*0*	*6*	*14*

Table 4: Percentage of studies achieving each index on the novel Quality of Therapy Reporting measure (QTR)

	N (of 93 studies)	% of studies
Method given	71	76%
Materials/items specified	37	40%
Service delivery clear	80	86%
Specified responses	48	52%
Info given when incorrect	46	49%

Mean total quality score = 2.94 (out of 5), SD 1.54

Table 5: Mean Quality of Therapy Reporting (QTR) score by design type (n= 93 studies)

Design	**Mean QTR/5**	**SD**
Randomised Controlled Trial	2.13	1.40
Parallel	2.74	1.42
Crossover	3.62	1.19
Case series	4.00	1.34

Figure 1: Literature Selection and Review Details

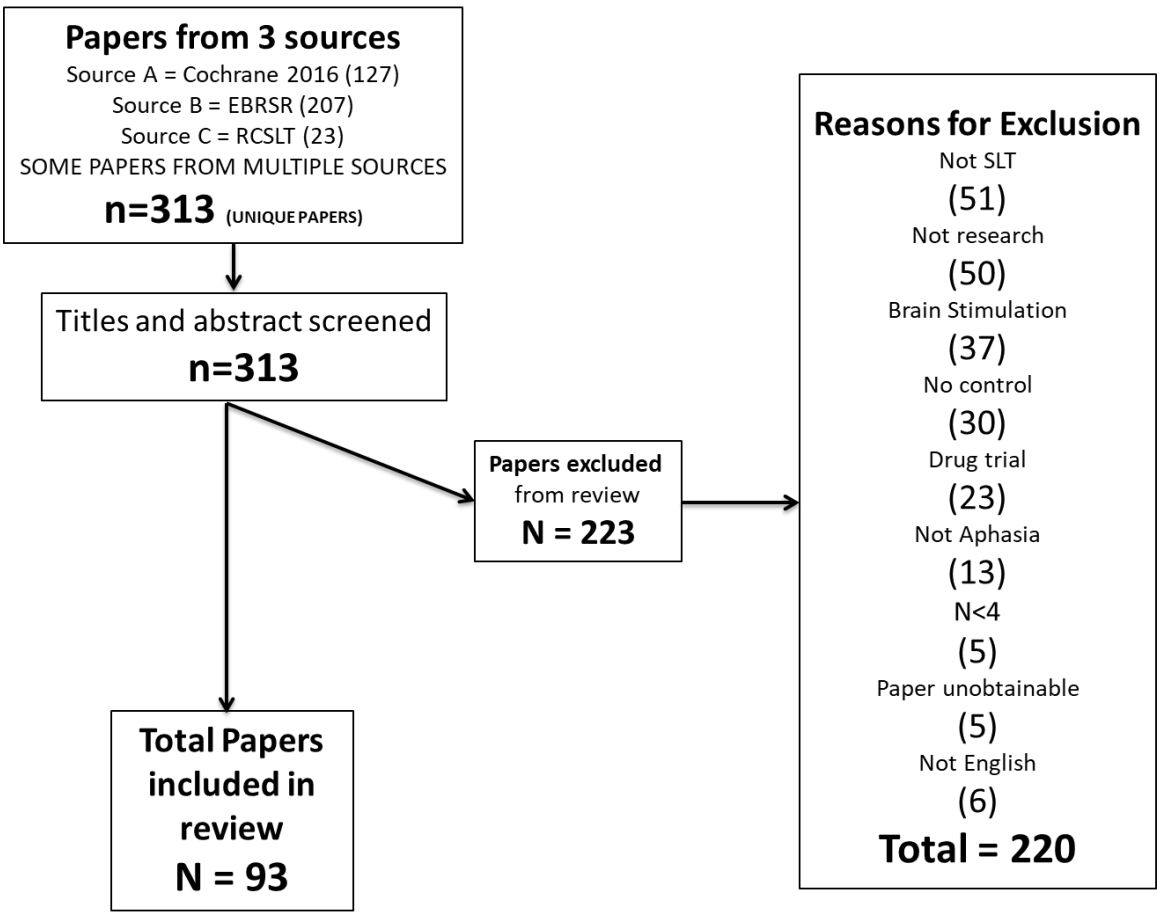

Figure 2: Proportions of treatment categories reported (n= 93 studies)

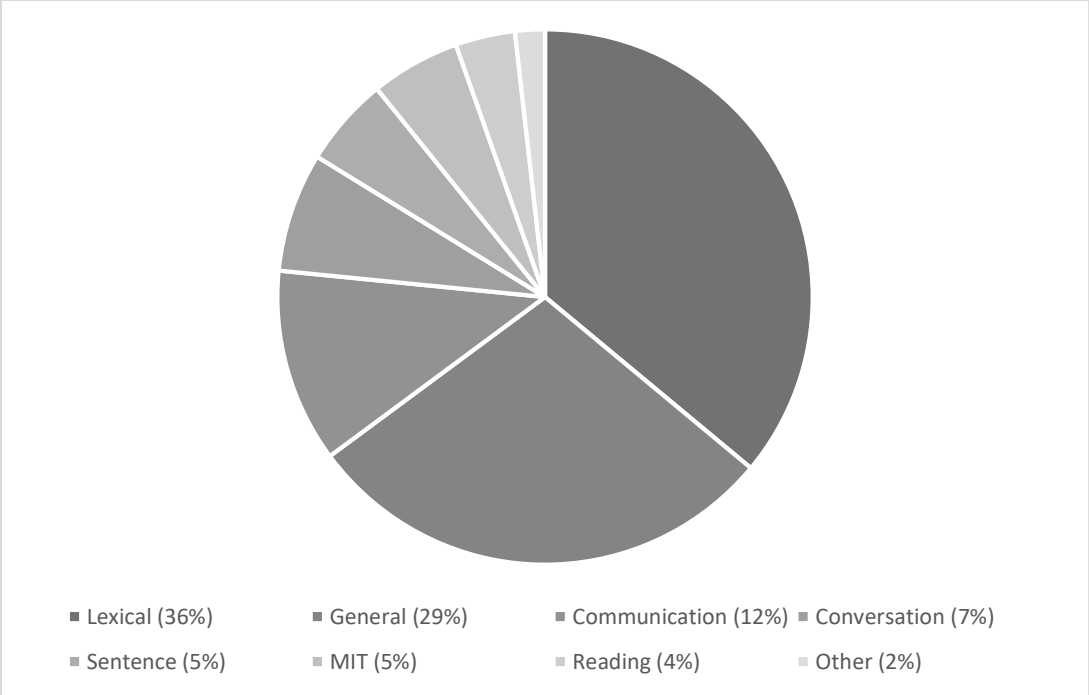

Figure 3: Number of lexical versus general treatments investigated across the four research designs (n=72 treatments)

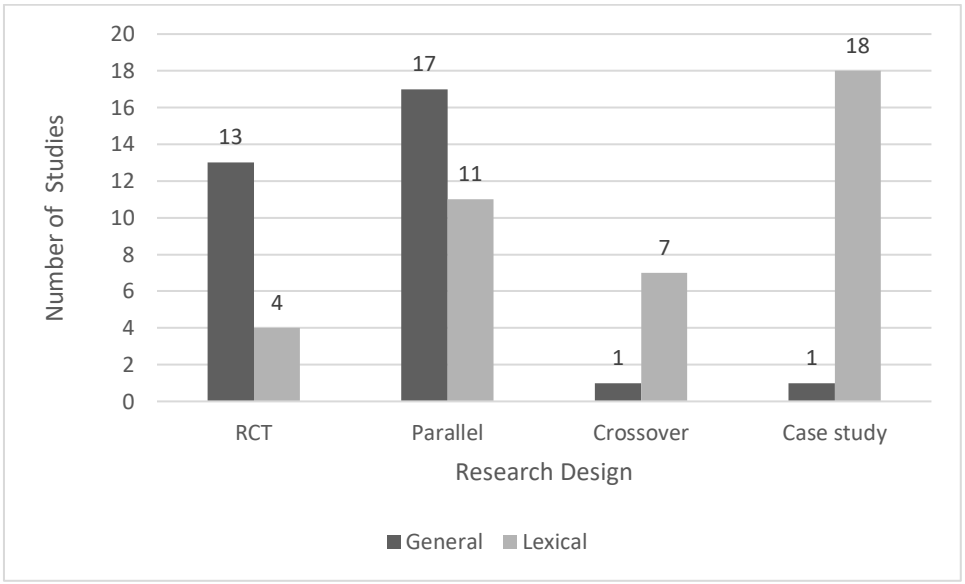
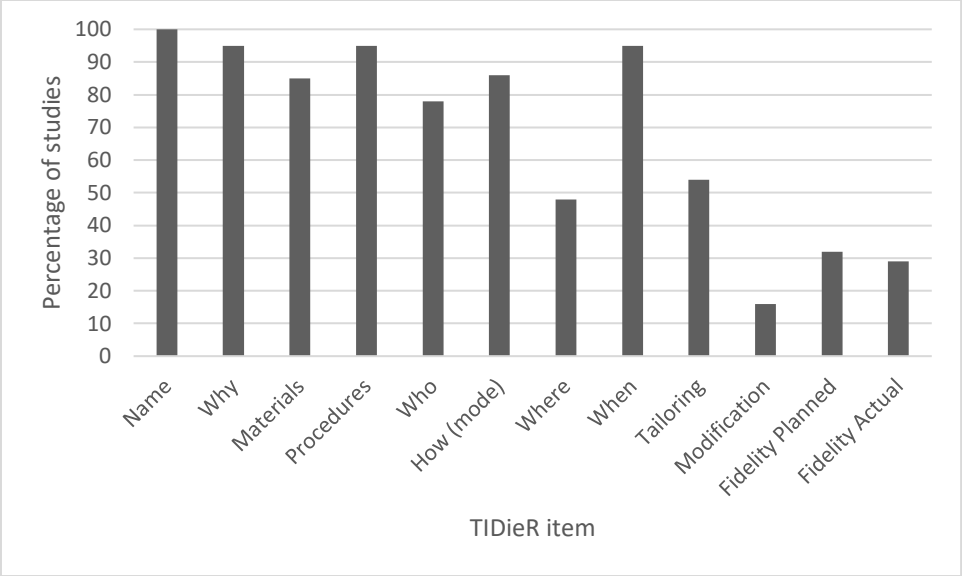
Figure 4 percentage of studies with information on each TIDieR checklist item (n=93 studies)