APHASIOLOGY
https://doi.org/10.1080/02687038.2022.2037502

Routledge
Taylor & Francis Group

Developing, monitoring, and reporting of fidelity in aphasia trials: core recommendations from the collaboration of aphasia trialists (CATs) trials for aphasia panel

Nicholas Behn[a], Madeleine Harrison[b], Marian C Brady[c], Caterina Breitenstein[d], Marcella Carragher[e], Julius Fridriksson[f], Erin Godecke[e,g], Argye Hillis[h], Helen Kelly[i], Rebecca Palmer[j], Miranda L. Rose[e], Shirley Thomas[k], Donna Tippett[h], Linda Worrall[l], Frank Becker[m,n] and Katerina Hilari[a]

[a]Centre for Language and Communication Science Research, School of Health Sciences, City, University of London, London, UK; [b]Department of Oncology and Metabolism, The University of Sheffield, Sheffield, UK; [c]Nursing,Midwifery and Allied Health Professions Research Unit, Glasgow Caledonian University, Glasgow UK; [d]Department of Neurology with Institute of Translational Neurology, University of Muenster, Muenster, Germany; [e]Centre of Research Excellence in Aphasia Recovery and Rehabilitation, La Trobe University, Melbourne Australia; [f]Department of Communication Sciences and Disorders, University of South Carolina, South Carolina, Columbia, USA; [g]Department of Speech Pathology, Edith Cowan University, and Sir Charles Gairdner Hospital, Perth, Australia; [h]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [i]Department of Speech & Hearing Sciences, University College Cork, Cork, Republic of Ireland; [j]Department of Health Service Research,School of Health and Related Research (ScHARR), The University of Sheffield, Sheffield, UK; [k]School of Medicine, University of Nottingham, Nottingham UK; [l]School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia; [m]Sunnaas Rehabilitation Hospital, Nesodden, Norway; [n]Institute of Clinical Medicine, University of Oslo, Oslo Norway

ABSTRACT

Background: Developing, monitoring, and reporting of fidelity are essential and integral components to the design of randomised controlled trials (RCTs) in stroke and aphasia. Treatment fidelity refers to the degree to which an intervention is delivered as intended and is directly related to the quality of the evidence generated by RCTs. Clear documentation of treatment fidelity in trials assists in the evaluation of the clinical implications of potential benefits attributed to the intervention. Consideration of the implementation requirements of a research-based intervention as intended in a clinical context is necessary to achieve similar outcomes for a clinical population. Despite this, treatment fidelity is rarely reported in RCTs of aphasia intervention.

Aim: To describe fidelity strategies and develop core recommendations for developing, monitoring, and reporting of fidelity in aphasia intervention RCTs.

Scope: Relevant conceptual frameworks were considered. The Behaviour Change Consortium comprehensive framework of fidelity was adopted. It includes five areas: study design, training providers, delivery of treatment, treatment receipt, and treatment enactment. We explored fidelity in RCTs with a range of complex aphasia interventions (e.g., ASK, Big CACTUS, COMPARE, FCET2EC, POLAR, SUPERB, and VERSE) and described how different trial design factors (e.g., phase of

ARTICLE HISTORY

Received 24 September 2021
Accepted 31 January 2022

KEYWORDS

Aphasia; trials; fidelity; adherence; recommendations

CONTACT Katerina Hilari K.Hilari@city.ac.uk City, University of London, London, UK

trial, explanatory vs. pragmatic, number and location of sites, and number and type of treatment providers) influenced the fidelity strategies chosen. Strategies were mapped onto the five areas of the fidelity framework with a detailed exploration of how fidelity criteria were developed, measured, and monitored throughout each trial. This information was synthesised into a set of core recommendations to guide aphasia researchers towards the adequate measurement, capture, and reporting of fidelity within future aphasia intervention studies.

Conclusions/Recommendations: Treatment fidelity should be a core consideration in planning an intervention trial, a concept that goes beyond treatment adherence alone. A range of strategies should be selected depending on the phase and design of the trial being undertaken and appropriate investment of time and costs should be considered.

This paper is part of a special series of papers in *Aphasiology* on Methodological Issues in Aphasia Trials. The series comprises tutorial-type papers with core recommendations for aphasia intervention studies and RCTs. The series is guest edited by Professor Katerina Hilari, Dr Caterina Breitenstein, Dr Erin Godecke, Dr Helen Kelly, and Professor Miranda Rose on behalf of the Trials for Aphasia Panel of the Collaboration of Aphasia Trialists https://www.aphasiatrials.org/

Introduction

Treatment fidelity refers to the "methodological strategies used to monitor and enhance the reliability and validity of behavioural interventions" (Bellg et al., 2004, p. 443) and is becoming increasingly recognised as an important quality marker for clinical trials (Bellg et al., 2004; Mowbray et al., 2003; Walker et al., 2017). Reporting fidelity can assist researchers to identify the essential elements that help make a treatment work (Carroll et al., 2007). Demonstrating fidelity is an important component in determining an intervention's replicability, efficacy, and effectiveness.

Fidelity is a multifaceted concept that has evolved over time (Bellg et al., 2004; Carroll et al., 2007). Early definitions of fidelity focused on *treatment delivery*, which Moncher and Prinz (1991) separated into issues related to treatment integrity and differentiation. Treatment integrity relates to whether a treatment is carried out as intended with adequate levels of adherence and competence (Perepletchikova et al., 2007); treatment differentiation refers to whether one treatment differs from another (Kazdin, 1986). Treatment integrity and, in particular, adherence to the protocol have historically been considered as the main factors demonstrating sufficient fidelity (Carroll et al., 2007; Mowbray et al., 2003; Orwin, 2000; Walker et al., 2017). *Adherence* to a treatment protocol is the minimum standard for fidelity, referring to the degree to which providers deliver the intervention as specified in the manual (Carroll et al., 2007; Perepletchikova et al., 2007). *Treatment receipt* and *enactment* were later added to expand further the definition of fidelity (Lichstein et al., 1994). Treatment receipt refers to a participant's ability to understand and perform the treatment skills (demonstrated, e.g., in pre-post tests, homework, and self-practice logs), while treatment enactment refers to a participant's ability to independently use learnt treatment skills in relevant daily-life settings (demonstrated, e.g., in direct observation, self-report, or follow-up discussion).

The Treatment Fidelity Workgroup of the National Institutes of Health Behaviour Change Consortium (BCC) reviewed existing definitions and treatment fidelity strategies and developed recommendations for consistent reporting (Bellg et al., 2004). The group identified two other areas related to fidelity: study design, which refers to methods that ensure a study can adequately test the hypotheses under investigation; and training providers, which are methods that ensure treatment providers are satisfactorily trained to deliver the treatment. These five areas of fidelity (*study design, training providers, delivery of treatment, receipt of treatment, and enactment of treatment skills*) will be used and referred to throughout this manuscript. The BCC workgroup has further described and refined methods for assessing, monitoring, and enhancing treatment fidelity for all five areas (Borelli et al., 2005; Borrelli, 2011).

Monitoring fidelity is important for maintaining the internal and external validity of a study (Moncher & Prinz, 1991). Importantly, in terms of internal validity, fidelity results may help explain negative, null, or ambiguous findings (Hohmann & Shear, 2002; Resnick et al., 2005). In the absence of fidelity, it is difficult to determine whether a null trial was driven by poor adherence to a specified protocol or ineffectiveness of the treatment itself (Carroll et al., 2007; Perepletchikova et al., 2007). Moreover, monitoring fidelity can help decrease deviations from the treatment protocol, including any instances of therapist or programme drift (Mowbray et al., 2003). Measuring drift is important to show provider skills are maintained consistently across the trial with no change or decay from the beginning to the end of treatment.

In terms of external validity, the benefits of considering and evaluating fidelity include an opportunity to identify or confirm the active ingredients of treatment that contribute to efficacy findings (Bellg et al., 2004; Moncher & Prinz, 1991). Active ingredients refer to intervention-specific components serving as key levers of change (Abry et al., 2015). Fidelity monitoring enables researchers to replicate the findings of the study and make comparisons across other similar studies, which is important for maintaining external validity. Finally, the lack of standardisation of provider training and monitoring of adherence to treatment protocols will impede study replication (Mowbray et al., 2003; Resnick et al., 2005).

When designing a fidelity evaluation, it is essential to consider the phase and purpose of the trial. In 2000, The Medical Research Council (Campbell et al., 2000) described a series of phases in the evaluation of complex interventions ranging from pre-clinical (i.e., laboratory animal studies) and early studies on the safety and modelling components (phase I) to preliminary evaluation needed to identify key components and test the feasibility of an intervention (phase II), to comparing the intervention to an appropriate alternative (phase III) and, lastly, testing whether an intervention works in everyday practice and gains are maintained in the long term (phase IV).

The purpose of the trial—explanatory or pragmatic—should also be considered. An explanatory trial determines the efficacy of an intervention in tightly controlled conditions to establish whether the intervention "can" work under ideal conditions (Zwarenstein et al., 2008). In contrast, pragmatic trials determine the effectiveness of an intervention if it were implemented in routine clinical practice, by asking "does it work in the 'real world'?"; the latter can often result in a more flexible application of the intervention with a more inclusive sample and multiple providers as well as individual adjustments of the therapeutic approach (Zwarenstein et al., 2008). Where a trial sits on the explanatory–pragmatic continuum can be determined using the PRagmatic Explanatory Continuum Indicator Summary 2 (PRECIS-2; Loudon et al., 2015).

The recommended practices for *ensuring* and *monitoring* fidelity proposed by the BCC are more applicable to explanatory than to pragmatic trial designs. Detailed recommendations for ensuring fidelity in pragmatic trials are yet to be published. However, as more pragmatic trials focus on interventions closer in nature to routine care evaluation, fidelity frameworks from the implementation literature, such as the Conceptual Framework for Implementation Fidelity (Carroll et al., 2007), have been utilised (Masterson-Algar et al., 2014; J. Palmer et al., 2019). In pragmatic, routine care trials, it is essential that strategies to enhance fidelity, such as provider training and relaying fidelity evaluation findings back to providers, do not go beyond that which can be delivered in routine clinical practice outside of a trial (Miller & Rollnick, 2014). Implementation fidelity models place greater focus on *measuring* the fidelity, with strategies to ensure fidelity needing to be built into the intervention itself, rather than relying on external trial processes (Carroll et al., 2007).

From development to reporting, fidelity should be an integral component of intervention trial design (Walker et al., 2017). Despite this, treatment fidelity procedures were not included within international clinical trial reporting guidelines such as the Consolidated Standards of Reporting Trials (CONSORT 2010, Eldridge et al., 2016) until the Template for Intervention Description and Replication (TIDieR) extension (Hoffman et al., 2014).

In stroke rehabilitation and aphasia studies, fidelity is rarely reported. For example, in published stroke rehabilitation trials during 2015, fidelity was reported for fewer than 10% of trials, with most studies addressing a single aspect of fidelity such as adherence (Walker et al., 2017). In post-stroke aphasia intervention studies, the measurement and reporting of fidelity is still in its infancy. A review of 149 studies published between 2002 and 2011 revealed that only 14% of studies reported on fidelity (Hinckley & Douglas, 2013). Systematic reviews of specific interventions for people with post-stroke aphasia have highlighted that fewer than 27% of studies reported fidelity (Cherney et al., 2008, 2013; Faroqi-Shah et al., 2010; Purdy et al., 2019; Rose et al., 2013; Simmons-Mackie et al., 2010, 2016). More recently, 21% of aphasia RCTs from 2012 to 2017 explicitly reported fidelity (Brogan et al., 2019). In these 42 aphasia RCTs, only one study addressed all five fidelity areas (Marshall et al., 2016). Approximately half the studies addressed either training providers or delivery of treatment (Brogan et al., 2019).

In designing a fidelity evaluation plan, criteria for fidelity assessment need to be clearly identified and specified a priori (Mowbray et al., 2003). This process involves specifying the active ingredients of a treatment and defining "standard operating procedures" to ensure the treatment is delivered as planned (Moncher & Prinz, 1991). A fidelity protocol should include the training and ongoing supervision of providers and development of a treatment manual to help minimise provider or therapist drift (Hildebrand et al., 2012; Moncher & Prinz, 1991). Further criteria to measure fidelity should also be established; these criteria may be structural (e.g., the planned frequency and intensity of contact), which require less subjective judgements, or process-related (e.g., staff–client or client–client interactions), which are more subjective and often based on observation (Mowbray et al., 2003). Direct observation by trained and reliable raters of treatment sessions using a prescribed list of criteria (*or behaviours*) is considered the gold standard of fidelity measurement (Hart, 2009; Kaderavek & Justice, 2010; Markowitz et al., 2000; Moncher & Prinz, 1991; Mowbray et al., 2003). In their review of aphasia studies, Hinckley and Douglas (2013) report that about half of the studies (13/21) checked adherence by having raters review 10–20% of videotapes of sessions for the presence of target behaviours.

The TIDieR extension to the CONSORT 2010 statement is recommended to encourage more transparent reporting of treatments. Not only does this template advocate for a complete and detailed description of an intervention (e.g., materials, procedure, how, where, when, and how much), but fidelity is also a core part. Two items (11 and 12) explicitly refer to planned and actual fidelity strategies undertaken in an intervention trial. These items have since been matched to the BCC's five areas of fidelity (Item 11: study design, training providers, delivery of treatment; Item 12: receipt of treatment, enactment of treatment skills). The actual outcome of planned fidelity strategies is less commonly reported in aphasia intervention trials (Brogan et al., 2019).

While fidelity monitoring and reporting may not have been historically common practice in aphasiology, more recently, the field has become increasingly aware of the importance of the development, implementation, evaluation, and reporting of fidelity strategies. This paper aims to provide examples of how fidelity was addressed across recent aphasia RCTs of different types and scales and build on existing frameworks to develop core recommendations for developing, monitoring, and reporting of fidelity.

Scope/Methods

To explore the measurement of fidelity and strategies used, seven RCTs for people with post-stroke aphasia from four countries (Australia, Germany, UK, USA) were examined. Considering the latest review of fidelity in aphasia RCTs covering trials up to 2017 (Brogan et al., 2019), we specifically included RCTs that were published or ongoing (with a published protocol) in the period from 2017 to the present. The trials needed to have a detailed description of planned fidelity as a minimum to be used as exemplars of current practice in fidelity monitoring and reporting. The selected studies comprised four explanatory (phases II and III) and three pragmatic (phases III and IV) trials. Explanatory trials included one feasibility and three definitive trials. The feasibility trial was of a peer-befriending intervention for people with aphasia and low levels of distress when discharged from hospital and active interventions cease (typically <6 months post-stroke) (SUPERB, Hilari et al., 2019). The three definitive trials were intensive aphasia therapy in the very early weeks post-stroke (VERSE, Godecke et al., 2021); multi-modal and constraint-induced aphasia therapy in the chronic phase (>6 months) post-stroke (COMPARE, Rose, Copland et al., 2019); and aphasia language therapy more than 12-months post-stroke (POLAR, Spell et al., 2020). Pragmatic trials included a psychosocial intervention to prevent depression in the early months post-stroke (ASK, Worrall et al., 2016), self-managed computerised speech and language therapy more than 4-months post-stroke (Big CACTUS, Palmer et al., 2020; R.2019), and intensive "agreed best practise" aphasia therapy starting at least 6-months post-stroke (FCET2EC, Breitenstein et al., 2017). Further information about trial design, sample size, and intervention characteristics is provided for each trial in Table 1.

To select studies for inclusion and describe the fidelity strategies of each trial, we considered peer-reviewed academic publications, information in the public domain including trial protocols, fidelity monitoring plans, and treatment manuals, and sought additional information from study authors. First, studies needed to have provided information on fidelity in either peer-reviewed publications and/or other forums in the public domain as above. Then, the study authors of each trial rated their own study and whether they had "met", "partially met",

Table 1. Study and intervention characteristics of aphasia intervention trials.

	POLAR	COMPARE	ASK	SUPERB	VERSE	FCET2EC	Big CACTUS
Study reports	Spell et al. (2020)	Rose et al. (2019); Rose et al. (2019)	Worrall et al. (2016); Carragher et al. (2019)	Hilari et al. (2019); Hilari et al. (2021); Behn et al. (2021)	Godecke et al. (2016); Godecke et al. (2021); under submission)	Baumgaertner et al. (2013); Breitenstein et al. (2017)	Palmer et al. (2020) Harrison (2019)
RCT phase and type*	Phase II, explanatory, multicentre, open-label, blinded endpoint, crossover RCT	Phase III, explanatory, multicentre, open-label, blinded endpoint, parallel group RCT	Phase III, pragmatic, superiority multicentre, cluster, double-blind, parallel group RCT	Phase II, explanatory, feasibility, multicentre, single-blind (intervention arm)/ double-blind (control arm), parallel group RCT	Phase III, explanatory, multicentre, open-label, three arm, blinded endpoint, parallel group RCT	Phase III/IV, pragmatic multicentre, open-label, blinded end point, parallel group RCT	Phase IV, pragmatic, superiority, multicentre, three arm, single-blind, parallel group RCT
Actual/target sample size	120 (target)	216 (actual)	344 (target)	56 (actual)	246 (actual)	158 (actual)	278 (actual)
Number of sites/ setting	2 university clinics/ community	Urban community facilities and University clinics	20 clusters (health regions)	5 Hospitals + linked community services; 13 GP practices	17 acute care hospital; 45 subacute hospital and home	19 inpatient and outpatient centres; day-care treatment facilities; hospitals; independent SLTs	21 community sites
Aphasia phase	≥12 months	>6 months	<6 months and study participation continued to 12-months post-onset	<6 months	<15 days	≥6 months	>4 months

(*Continued*)

Table 1. (Continued).

	POLAR	COMPARE	ASK	SUPERB	VERSE	FCET2EC	Big CACTUS
Intervention and control arms	2 arms: semantically focused therapy using 3 approaches vs. phonologicallyfocused therapy using 3 approaches	3 arms: usual care CIAT Plus M-MAT	2 arms: ASK psychosocial intervention vs. attention control (secondary stroke prevention information)	2 arms: usual care + peer-befriending vs. usual care control	3 arms: usual Care + 15–20 hours of standard ward-based aphasia therapy vs. usual care + 15–20 hours prescribed therapy (targeting spoken output) vs. usual care control	2 arms: intensive agreed best practice + self-managed (computerised) language exercises vs. waiting list control	3 arms: usual care + self-managed computerised speech and language therapy vs. usual care + attention control (puzzle books and phone calls) vs. usual care control
Provider	Speech and language therapist	Speech and language therapist	Speech and language therapist	Person with aphasia	Speech and language therapist	Speech and language therapist	Self-managed with support from speech and language therapist and volunteer/ assistant

*See glossary of terms

Note. POLAR = Predicting Outcomes of Language Rehabilitation; COMPARE = Constraint-induced Or Multi-modal Personalised Aphasia Rehabilitation; ASK = Action Success Knowledge; SUPERB = SUpporting wellbeing through PEeR Befriending; VERSE = Very Early Rehabilitation in SpEech; FCET2EC = From Controlled Experimental Trial to Everyday Communication; CACTUS = Clinical and cost-effectiveness of Aphasia Computer Treatment vs. Usual Stimulation or attention control long-term post-stroke; CIAT-Plus = Constraint-Induced Aphasia Therapy Plus; M-MAT = Multi-modality Aphasia Therapy.

or "not met" the fidelity practice requirements according to treatment fidelity strategies within the five areas of fidelity proposed by Bellg et al. (2004): study design, training providers, delivery of treatment, receipt of treatment, and enactment of treatment skills. The study authors' ratings were reviewed for inclusion of each study in this review by the first and last authors of the present paper to ensure fidelity strategies were used across at least four of the five areas.

Each trial was then rated by two different raters, who were experienced aphasia trialists and co-authors of this paper but who were independent of the RCT study they were rating. Raters reviewed the published documents and rated whether fidelity strategies were met in each RCT. The independent raters were not aware of the ratings provided by the study authors. Planned (if ongoing) and actual (if completed) fidelity strategies were rated. Each rater was required to independently review the published documents for fidelity strategies relevant to the specific area and make a judgement as to whether the trial had "met", "partially met", or "not met" the requirements. Resources to rate fidelity, that is, relevant papers (Bellg et al., 2004; Borelli et al., 2005; Borrelli, 2011) were given to the raters; but no training was provided.

To examine the inter-rater reliability of the independent raters, Gwet's AC1 was used (Gwet, 2002, 2014). Unlike Cohen's Kappa, Gwet's AC1 can be calculated even when one rater's scores are a constant/do not vary; Gwet's AC1 is less susceptible to skew due to prevalence (Wongpakaran et al., 2013). A reliability coefficient of 0.81–1.00 is considered very good; 0.61–0.80, good; 0.41–0.60, moderate; 0.21–0.40, fair; and below 0.20, poor (Gwet, 2014). Statistical analyses were carried out using STATA version 15.

As disagreements in some of the ratings emerged, definitions of fidelity areas and fidelity strategies were discussed among the raters assigned to the same study to explore differences. Raters then met to reach an agreement on the rating by consensus. Based on the review of published documents for each trial, fidelity ratings, and comments on the seven trials by authors and raters, key fidelity strategies used across trials were described. Through a review of these strategies (planned or actual), the author team then agreed on a set of recommendations for which areas of fidelity should be considered essential or desirable across trial phases and the explanatory to pragmatic continuum.

Results

For inter-rater reliability of fidelity ratings, there was a fair to very good level of agreement between raters. There was fair agreement for the ASK (Gwets AC1 = 0.33 [95% CI, −0.10, 0.75], $p = 0.12$) and VERSE (Gwets AC1 = 0.37 [95% CI, −0.04,0.79], $p = 0.07$) trials. There was moderate agreement for COMPARE (Gwets AC1 = 0.42 [95% CI, 0.20, 0.82], $p = 0.04$), Big CACTUS (Gwets AC1 = 0.51 [95% CI, 0.13, 0.89], $p = 0.01$), and SUPERB trial (Gwets AC1 = 0.57 [95% CI 0.20, 0.95], $p = 0.005$). There was very good agreement for the POLAR (Gwets AC1 = 0.84 [95% CI, 0.60, 1.00] $p < 0.001$) and FCET2EC (Gwets AC1 = 0.86 [95% CI, 0.63, 1.00], $p < 0.001$) trials. Overall, across all raters and trials, there was 63% agreement (ranging from 43.8% to 87.5% agreement). However, when the categories "partially met" and "met" were collapsed into one, the agreement increased to 84% (ranging from 75% to 100%).

Table 2 shows the final ratings agreed for each area of fidelity and the extent to which each component was "met", "partially met", or "not met". Training providers and delivery of treatment were addressed more completely than other areas of fidelity. Treatment receipt and enactment were addressed less well.

There were subtle and distinct differences across all trials as to the strategies employed for each fidelity area according to the type and phase of the trial. Below is a summary of key strategies used across the seven trials (Table 3). It should be noted that fidelity strategies were either *planned* for ongoing trials at the time of writing this manuscript (i.e., POLAR, ASK, and COMPARE) or *actual reported* for completed trials (i.e., SUPERB, VERSE, FCET2EC, and Big CACTUS). It is noted that as Big CACTUS was a pragmatic trial, which aimed to *monitor* rather than ensure fidelity, many areas are marked as partially met, given the focus of the Bellg framework on ensuring fidelity. Instances of extended practice are identified to highlight examples where a trial went beyond the expectations of fidelity monitoring detailed by Bellg et al. (2004).

Strategies to address study design

In planning a treatment trial, the theory underpinning the treatment should be described and essential components or active ingredients identified. While this was completed across all trials, some trials did this more explicitly in publications focused on treatment fidelity (e.g., ASK, SUPERB, and VERSE). A number of trials also did pilot studies to explore the active ingredients in depth (e.g., ASK, Big CACTUS, COMPARE, SUPERB, and VERSE).

To ensure fidelity, key issues to consider when planning a study are dose within and across conditions and strategies to address implementation setbacks (Bellg et al., 2004). To ensure and monitor the same treatment dose *within* conditions, all trials planned to or recorded the length and number of all contacts. For trials with an independent practice component in the intervention arm, information was planned or recorded on the number and length of practice sessions (e.g., Big CACTUS, COMPARE, and FCET2EC). An example of extended practice in monitoring dose was in the VERSE trial, where participants in all trial arms had a proportion of treatment sessions videorecorded and independently assessed for overall treatment protocol adherence.

The data recorded on length and number of contacts were also used to ensure equivalence of dose *across* conditions. Other strategies were also employed, such as recording any speech and language therapy and other services received by participants including those in the control arm (e.g., ASK, Big CACTUS, COMPARE, FCET2EC, SUPERB, and VERSE); and having participants document therapy activity using a study-specific diary in the control arm or have providers keep a daily therapy log for those in the intervention arm (e.g., COMPARE and FCET2EC). In some cases, ensuring the same treatment dose within and equivalence of dose across conditions is not possible or desirable. For example, pragmatic trials in routine care, where there may be variability in how much therapy provision is funded by healthcare providers (e.g., FCET2EC) or for self-managed interventions where the aim is to increase the dose of therapy in the intervention arm only (e.g., Big CACTUS, and VERSE).

Table 2. Description of fidelity strategies and number of strategies rated as met, partially met, or not met in each trial.

	Planned fidelity			Actual fidelity			
	POLAR	ASK	COMPARE	SUPERB	VERSE	FCET2EC	Big CACTUS
Fidelity report	Spell et al. d(2020)	Carragher et al. (2019)	Rose et al. (2019)	Behn et al. (2021)	Godecke et al. (under submission)	Breitenstein et al. (2017)	Palmer et al. (2020); Harrison (2019)
Study design							
Ensure same treatment dose within conditions	Met	Partially met	Met	Met	Met	Partially met*	Not met
Ensure equivalent dose across conditions	Met	Partially met	Partially met	Partially met	Not met	Met	Not met
Plan for implementation setbacks	Partially met	Partially met	Partially met	Met	Met	Partially met	Partially met
Training providers							
Standardise training	Met	Met	Met	Met	Partially met	Met	Partially met
Ensure provider skill acquisition	Met	Met	Met	Met	Partially met	Met	Partially met
Minimise therapist drift	Met	Partially met	Partially met	Met	Met	Met	Partially met
Accommodate provider differences	Partially met	Met	Partially met	Met	Partially met	Met	Partially met
Delivery of treatment							
Control for provider differences	Met	Met	Met	Met	Met	Partially met	Partially met
Reduce differences within treatment	Met	Met	Met	Met	Met	Met	Partially met
Ensure adherence to protocol	Met	Met	Met	Met	Met	Met	Partially met
Minimise contamination between conditions	Met	Met	Met	Met	Met	Met	Partially met
Receipt of treatment							
Ensure participant comprehension	Met	Met	Partially met	Met	Met	Met	Partially met
Ensure participant ability to use cognitive skills	Met	Not met	Not met	Met	Partially met	Met	Partially met
Ensure participants ability to perform behavioural skills	Met	Not met	Met	Partially met	Partially met	Met	Partially met
Treatment enactment							
Ensure participant use of cognitive skills	Not met	Not met	Partially met	Partially met	Not met	Partially met	Met
Ensure participant use of behavioural skills	Not met	Not met	Partially met	Partially met	Not met	Met	Met

* In FCET2EC, the same treatment dose across participants was not an option as it was an ethical requirement to provide the maximum amount of therapy funded by the individual's health insurance plan. For the first (and most relevant) follow-up assessment after three weeks though, dose was highly comparable across participants within the intervention condition (at least 80% of the prespecified dose).

POLAR = Predicting Outcomes of Language Rehabilitation; COMPARE = Constraint-induced Or Multi-modal Personalised Aphasia Rehabilitation; ASK = Action Success Knowledge; SUPERB = SUpporting wellbeing through PEeR Befriending; VERSE = Very Early Rehabilitation in SpEech; FCET2EC = From Controlled Experimental Trial 2 Everyday Communication; CACTUS = Clinical and cost-effectiveness of Aphasia Computer Treatment vs. Usual Stimulation or attention control long-term post-stroke.

Table 3. Example strategies for each area of fidelity across recent aphasia RCTs.

	Rationale	Example strategies
Study Design	Ensures a study can adequately test its hypotheses in relation to underlying theory and clinical processes	• Specified theory underpinning treatment and essential components identified. • Record the length, number, and frequency of all contacts including practice and homework sessions (in all trial arms) • Videorecorded sessions to independently assess for adherence • Study-specific record sheet or daily therapy logs • Supervision and training of providers • Train more providers than required • Allow additional time for recruitment of participants and intervention to manage setbacks
Training providers	Ensures treatment providers have been satisfactorily trained to deliver intervention	• Same or certified trainers with set qualifications and/or experience levels • Training material described • Standardised training • Length of training recorded • Videorecorded training and/or treatment sessions to independently assess providers skills and competence • Feedback to providers • Knowledge quiz to providers at regular intervals • Regular provider face-to-face supervision, site visits, or telephone calls • Booster or refresher training or support • Predetermined acceptable level of protocol adherence for reported treatment sessions
Delivery of treatment	Monitor and improve the delivery of the intervention so that that the protocol is delivered as intended	• Daily therapy logs or visit records • Monitor information recorded by computer software • Standardisation of intervention elements to participants • Recording adverse events, complaints, problems, or protocol deviations • Videorecorded intervention sessions and independently assess for adherence with fidelity checklist • Audio or videorecorded sessions • Qualitative participant interviews • Participant questionnaires of their experiences • Supervision and feedback to providers at regular intervals or when required • Standardised intervention manual and/or scripted intervention protocol • Providers deliver a single intervention • Providers do not liaise with other conditions or trial arm • Different intervention conditions occur in different locations • Close monitoring of intervention or services received by all trial arms • Study participants in control arm explicitly not aware of intervention (if feasible)

(*Continued*)

Table 3. (Continued).

	Rationale	Example strategies
Treatment receipt	Monitor and improve the ability of participants to understand the intervention and perform the skills being taught in intervention sessions (behavioural and cognitive)	• Providers have skills or been trained to facilitate communication • Inclusion and exclusion criteria ensure participants have skills to participate • Incremental or stage structure to learning skills • Review of videotaped intervention sessions, daily therapy logs, or homework exercises • Ratings of goal achievement • Monitoring functional tasks, e.g., conversation and group performance
Treatment enactment	Monitor and improve the ability of participants to perform the skills being taught in real-life settings	• Qualitative participant interviews • Opportunities to practice trained items/behaviours in functional communication tasks, e.g., conversation and self-reported outcomes or behavioural measures

To manage setbacks, trials described careful ongoing monitoring through supervision and training (e.g., ASK, COMPARE, FCET2EC, POLAR, and SUPERB), including training more providers than were required (e.g., FCET2EC and VERSE) and allowing additional time to stagger recruitment of participants, training, and/or completion of intervention (e.g., Big CACTUS, COMPARE, SUPERB, and VERSE).

Strategies to address training providers

Training providers refers to the training of treatment providers to ensure that they can adequately and consistently deliver the intervention. Standardisation of training is a strategy described in all seven trials and includes the use of the same and/or certified trainers, set training materials, and/or a clear description of the length of training. An example of extended practice was having the training sessions videotaped and independently rated against a fidelity checklist (e.g., SUPERB).

To demonstrate that researchers ensured or monitored the acquisition of skills by providers and accommodated for differences between them, provider training included opportunities for discussion and questions. Specific strategies included video recording to ascertain a provider's understanding of the training content followed by monitoring of therapy sessions, which were also recorded (e.g., COMPARE), review of video recording for each provider's first intervention sessions (e.g., ASK) or a selection of treatment sessions (e.g., VERSE), feedback given after a session delivered in the first week of treatment (e.g., POLAR), independent rating of videotaped training sessions for skill acquisition (e.g., SUPERB), or a quiz to evaluate the knowledge of providers (e.g., Big CACTUS).

To minimise provider/therapist drift, a range of strategies was used, including regular, scheduled supervision (e.g., FCET2EC, POLAR, and SUPERB) or refresher supervision and training as necessary (e.g., ASK, COMPARE, and VERSE). To monitor drift, strategies included monitoring written therapy documentation, which underwent quality assurance checks and at least weekly telephone conversations with providers (e.g., FCET2EC), or giving providers a quiz at 5, 10, and 15 months post-training to check provider knowledge and what they had understood, internalised, and were applying during the trial (e.g., Big CACTUS). Further examples involved a review of videotaped intervention sessions to give feedback to providers to ensure skill acquisition (e.g., ASK, COMPARE, POLAR, SUPERB, and VERSE) and regular site visits (e.g., VERSE).

To accommodate for provider differences, training was given to all providers regardless of existing qualifications and/or experience levels (e.g., Big CACTUS, COMPARE, FCET2EC, and VERSE). Small group training was reported to ensure the individual needs of providers, who were people with aphasia, were met (e.g., SUPERB). In addition, a predetermined clinically acceptable level of protocol adherence for treatment sessions was reported to ensure all providers were consistent in their delivery (e.g., ASK, COMPARE, VERSE, and FCET2EC).

Strategies to address delivery of treatment

Delivery of treatment refers to strategies that determine if the intervention is delivered as intended. To control for provider differences, how the treatment was delivered by individual providers was monitored in different ways, including daily therapy logs or

visit records (e.g., COMPARE, SUPERB, ASK, FCET2EC, and VERSE), documentation of therapy choice (e.g., Big CACTUS), standardisation of training for key therapeutic elements such as the rate of presentation of materials and therapeutic techniques (e.g., COMPARE, POLAR, and FCET2EC), and recording adverse events, complaints, and/or problems reported (e.g., COMPARE, FCET2EC, SUPERB, VERSE, and ASK). Ratings of session adherence help to control for differences between treatment providers (e.g., ASK, COMPARE, POLAR, SUPERB, and VERSE). In SUPERB, where the providers of the peer-befriending intervention were people with aphasia, criteria for matching a provider with a participant were monitored and recorded. Examples of extended practice included qualitative interviews for a selection of participants at the end of the trial to assess provider perceptions and experiences of the intervention (e.g., SUPERB, VERSE, and ASK) or completion of a questionnaire by providers to describe differences in the therapeutic bond, and goal and task agreement (e.g., Big CACTUS).

To reduce differences within treatment, strategies included training on key therapeutic elements (e.g., COMPARE and VERSE), self or independent ratings of session adherence (e.g., ASK, COMPARE, POLAR, SUPERB, VERSE), supervision and feedback to providers (e.g., ASK, COMPARE, FCET2EC, POLAR, SUPERB, and VERSE), use of a standardised treatment manual (e.g., Big CACTUS, FCET2EC, SUPERB, and VERSE), and use of a scripted and continuously monitored intervention protocol (e.g., FCET2EC).

To ensure and monitor adherence to protocols, intervention sessions were monitored using videotaped sessions rated against fidelity checklists that contained key criteria (e.g., ASK, COMPARE, SUPERB, and VERSE) or using a combination of face-to-face and videotaped sessions (e.g., POLAR). These ratings were completed by an independent assessor in some studies (e.g., SUPERB, VERSE, and COMPARE). The proportion of face-to-face or videotaped intervention sessions rated overall ranged from 10% to 25% (e.g., ASK, COMPARE, POLAR, SUPERB, and VERSE) with two trials intending to rate all first treatment sessions (e.g., ASK and COMPARE). Information about the intervention dose was also recorded and monitored, including the number and length of intervention sessions with protocol deviations (e.g., ASK, COMPARE, FCET2EC, SUPERB, and VERSE); daily therapy logs or visit records (e.g., COMPARE, FCET2EC, SUPERB, VERSE, and ASK); and monitoring of information recorded by therapy software (e.g., Big CACTUS). In some trials, feedback was given to the providers about adherence during the course of the trial (e.g., ASK, COMPARE, FCET2EC, SUPERB, and VERSE).

A range of strategies were employed to minimise contamination between conditions. Providers were trained to not liaise with providers of other conditions (e.g., VERSE) and to give a single intervention (e.g., ASK and SUPERB). Each healthcare/community location was responsible for a specific intervention condition (e.g., COMPARE and ASK); study participants in the control arm were not explicitly aware of the intervention tested (e.g., SUPERB and ASK). In two trials, contamination was not expected as no treatment was given by the provider during the waiting period (e.g., FCET2EC) or was limited due to the requirements of the computer software used in the treatment (e.g., Big CACTUS). Extended practice included monitoring of all trial conditions to determine what intervention they received (e.g., Big CACTUS, COMPARE, FCET2EC, and SUPERB). In one trial, a planned review of videotaped sessions from all conditions to monitor the risk of contamination was planned but partially achieved (e.g., VERSE). Should contamination be found, a protocol deviation would be logged and provider retraining commenced (e.g., COMPARE).

Strategies to address treatment receipt

Treatment receipt assesses whether the participant can comprehend the intervention and use the skills being taught. Strategies in this area included more subtle methods, such as training the providers or having professionals (i.e., speech and language therapists) who already have skills in supporting communication to deliver the intervention (e.g., ASK, Big CACTUS, FCET2EC, POLAR, SUPERB, and VERSE) or specifying minimum comprehension or cognitive ability as a study inclusion criterion to ensure sufficient skills to take part (e.g., Big CACTUS, FCET2EC, and VERSE). The design of some interventions included incremental (or adaptive) levels or a stage structure whereby a participant needed to comprehend the intervention and use the cognitive skills being taught to be able to progress (e.g., Big CACTUS, COMPARE, FCET2EC, and VERSE). An example of extended practice was the review of videotaped intervention visits or treatment logs to monitor the degree with which a provider ensured a participant's comprehension and had sufficient cognitive skills to carry out and benefit from the intervention (e.g., POLAR and SUPERB). In addition, providers reviewed completed homework exercises and monitored participant performance daily to ensure the participants had understood training instructions (e.g., FCET2EC).

To ensure a participant could perform the behavioural skills being taught, a range of strategies were used, including goal achievement (e.g., VERSE), functional communication ability measured by video recorded conversations (e.g., Big CACTUS), performance in group therapy sessions as monitored by therapy supervisors (e.g., FCET2EC), and an activity log that demonstrated the use of therapy skills completed between sessions (e.g., COMPARE).

Strategies to address treatment enactment

Treatment enactment evaluates whether the participant demonstrates the ability to "use the intervention" in real-life settings. "Using the intervention" means employing the cognitive and behavioural skills that were taught during the intervention. Compared to other fidelity areas, treatment enactment may be harder to capture and has received less attention in aphasia RCTs. Examples of strategies to use here include qualitative interviews of a purposive sample of participants to explore their experiences and the impact of the intervention on their performance (e.g., Big CACTUS and SUPERB), examining the extent with which the language learned was then used in everyday conversation with volunteers (e.g., Big CACTUS) or the use of self-reported outcomes or behavioural measures that probed the participants' use of the taught skills (e.g., simulated everyday life communication scenarios such as role-plays in FCET2EC and functional communication measures in COMPARE).

Recommendations

Having reviewed the literature on fidelity standards (Bellg et al., 2004; Borelli et al., 2005; Borrelli, 2011; Carroll et al., 2007) and fidelity strategies employed in a sample of ongoing and recently completed aphasia RCTs, we propose a set of recommendations for aphasia researchers to appropriately address fidelity when planning, conducting, and reporting aphasia intervention trials.

- Fidelity monitoring should be seen as an integral part of all intervention trials (whether explanatory or pragmatic).
- Intervention trials should have a fidelity monitoring plan/protocol at the outset of the trial (Feely et al., 2018).
- Explanatory trials need to *ensure* fidelity; pragmatic trials need to *monitor* fidelity and statistically control for any effects of variable fidelity. Depending on the phase of a trial and the question of interest, some areas of fidelity need to be addressed to a greater or lesser extent. For example, treatment enactment may be a lower priority in early *feasibility* trials that do not evaluate efficacy and effectiveness.
- *Study design*: The intervention dose within and between all trial arms should be ensured/monitored including the control group. Active ingredients or essential components of the intervention should be clearly defined, described, and operationalised in a fidelity checklist (or list of protocol behaviours to be observed).
- *Fidelity checklists* should be created systematically, including developing an intervention framework, obtaining stakeholder feedback, and piloting the checklist to improve inter-rater reliability (Walton et al., 2020).
- *Training providers*: a detailed description is required regarding how providers are trained before the commencement of intervention. Providers should be clear about the active ingredients of the intervention (e.g., fidelity checklists should be shared with providers). Supervision and feedback to providers during the intervention should be specified to monitor adherence to the protocol.
- *Delivery of treatment*: adherence to intervention should be ensured or monitored using a range of data collection strategies including written data (e.g., daily records and activity logs) and/or observation of sessions (e.g., videorecorded sessions). Sessions should be rated for adherence. Researchers should aim to rate adherence for a high proportion of sessions, as feasible within their trial design. Rating 20% of sessions is considered optimal (Borrelli, 2011); when this is not feasible, we would recommend rating a minimum of 10% of sessions. A predetermined acceptable level of session adherence should be set (e.g., 100% of a specific set of behaviours; 80% of the general target behaviours). Failure to reach an acceptable level should be recorded as a protocol deviation and a management plan to mitigate future instances should be described.
- *Treatment receipt*: participants' grasp of the information provided and ability to apply the trained behaviours should be assessed.
- *Treatment enactment*: participants' application of the behaviours/skills to real-life settings should be assessed. Assessment methods may include questionnaires, self-reports, follow-up interviews, and telephone calls (Bellg et al., 2004; Borrelli, 2011) as well as conversation samples or assessment of target performances in non-therapeutic environments.

Discussion

The aim of this paper was to describe how fidelity has been addressed across recent aphasia RCTs and to develop core recommendations for fidelity in aphasia intervention trials. Historically, fidelity monitoring was not consistently embedded within aphasia intervention studies. Previous reviews of fidelity monitoring in aphasia intervention studies have found that less than half explicitly assessed adherence via video recordings

of intervention sessions (Hinckley & Douglas, 2013), less than a quarter reported supervision of the intervention (Hinckley & Douglas, 2013) and in aphasia RCTs published prior to 2017 only around 50% reported on training providers and delivery of treatment (Brogan et al., 2019). The current paper demonstrates a shift in the field of aphasiology towards more consistent and comprehensive fidelity monitoring and reporting. Seven aphasia RCTs were reviewed, including trials of different phases (phases II to IV) and both explanatory and pragmatic trials. Encouragingly, four trials at least partially met all five areas of fidelity, with the remaining three trials at least partially meeting four areas. Such a finding indicates that fidelity is increasingly being considered in the planning and reporting of aphasia intervention trials. Furthermore, the interpretation of fidelity in aphasia RCTs has moved beyond treatment adherence, indicating progress from the historic focus of fidelity in stroke rehabilitation research (Walker et al., 2017). Based on the literature and the reviewed RCTs, we provided a detailed but not exhaustive description of potential fidelity strategies to use and developed initial recommendations to help guide researchers as to what areas of treatment fidelity to consider for a range of trial phases.

Pragmatic trials present a challenge to the Bellg et al. (2004) framework as they *monitor* rather than *ensure* fidelity. In such trials, firstly, any modifications to the intervention should be systematically documented, using, e.g., FRAME (Stirman et al., 2019), to understand their impact. Secondly, alternative frameworks of fidelity may also be appropriate such as that proposed by Carroll et al. (2007), which focuses on treatment adherence (i.e., content and dose) and the factors that may influence the degree with which the intervention is implemented in practice (i.e., comprehensiveness of policy description, strategies to facilitate implementation, quality of delivery, and participant responsiveness). However, the Carroll et al. framework appears to place less emphasis on the receipt and enactment of the treatment and conceptualises these areas as more related to participant responsiveness. To guide researchers in such instances, the fidelity strategies employed by two pragmatic trials included in this paper, i.e., FCET2EC and Big CACTUS, may be helpful.

Treatment receipt and enactment have been addressed in less than 12% of aphasia intervention trials (Brogan et al., 2019). All seven trials reviewed here at least partially reported on an element of treatment receipt and four reported on an element of treatment enactment. Both these areas of fidelity extend the focus beyond the provider to the participant, underlining that an intervention should not merely be provided, but also received and consecutively lead to a change in everyday life. These areas partially overlap with the assessment of treatment effectiveness, which is likely assessed beyond a trial's fidelity plan. The fidelity domain of enactment continues to present challenges for definition, assessment, and reporting. Enactment may be difficult to assess because of the complex nature of communication (not a single behaviour to observe), the heterogeneity of aphasia, and because of limited or absent carer support to facilitate the use of taught skills in the home and community settings. Expert consensus is needed to determine whether other assessments completed during a trial (e.g., activity logs, conversation samples, and talk time) may be suitable strategies to measure enactment.

A challenge within this paper related to achieving agreement between raters in whether a requirement was met, partially met, or not met; with reliability coefficients ranging from fair to very good. A large proportion of disagreement related to whether a requirement was met or partially met. One difficulty may lie in the strategies proposed by Borelli et al. (2005) and Borrelli (2011), which were designed for health behaviour change trials where one specific behaviour is targeted (e.g., smoking). In aphasia, typically a range of behaviours are addressed as part of the intervention. However, there was also inconsistency in the raters' interpretation of the criteria and strategies. Pilot rating of a study with discussion between assessors on criteria and rating discrepancies may help ensure a common understanding of criteria and higher agreement in the future.

Moreover, the recommendations proposed in the current paper were generated through discussion and consensus agreement within a group of experienced aphasia trialists interested in treatment fidelity. A systematic methodology such as an e-Delphi exercise may be considered in future research in this area to determine the most important aspects of fidelity assessment in aphasia rehabilitation. The focus of this paper has been on treatment fidelity, which is only one aspect of fidelity. Researchers may also want to consider fidelity of assessment, which has been reported for two of the trials comprised in the current paper (i.e., COMPARE, POLAR).

Conclusions

This paper describes how the field of aphasiology has progressed in its monitoring and reporting of treatment fidelity. Referring to a selection of ongoing or completed aphasia trials as exemplars, we highlight examples of fidelity strategies that represent good practice to guide planning of future stroke and aphasia intervention trials. Treatment fidelity should be considered as integral to trial designs in the evaluation of interventions. A range of strategies should be selected depending on the phase and design of the trial. Enhanced fidelity monitoring and reporting will support transparency of trial reporting, increase our confidence in the trial results, and ultimately improve clinical decision-making for the benefit of people living with aphasia.

Acknowledgments

ASK was funded by a grant from the National Health and Medical Research Council Grant (APP1060673); Big CACTUS was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme (12/21/01) and Tavistock Trust for Aphasia; COMPARE was funded by a National Health and Medical Research Council Grant (APP1083010); FCET2EC was funded by the German Federal Ministry of Education and Research (BMBF; identifier, 01GY1144) and the German Society for Aphasia Research and Treatment (GAB); POLAR was funded by National Institute on Deafness and Other Communication Disorders Grants P50 DC014664; SUPERB was funded by The Stroke Association, Psychological Consequences of Stroke – Priority Programme Award (PPA2015-03); VERSE was funded by the National Health and Medical Research Council Grant (APP1044973), The Tavistock Trust for Aphasia, and Edith Cowan University, Australia.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Nicholas Behn http://orcid.org/0000-0001-9356-9957
Marian C Brady http://orcid.org/0000-0002-4589-7021
Caterina Breitenstein http://orcid.org/0000-0002-6408-873X
Marcella Carragher http://orcid.org/0000-0002-7200-6968
Erin Godecke http://orcid.org/0000-0002-7210-1295
Argye Hillis http://orcid.org/0000-0002-5192-1171
Helen Kelly http://orcid.org/0000-0003-3694-2086
Rebecca Palmer http://orcid.org/0000-0002-2335-7104
Shirley Thomas http://orcid.org/0000-0003-0704-9387
Linda Worrall http://orcid.org/0000-0002-3283-7038
Frank Becker http://orcid.org/0000-0002-0857-0628
Katerina Hilari http://orcid.org/0000-0003-2091-4849

Glossary of terms

https://www.aphasiatrials.org/design-planning/

Trial design

Parallel group: where participants are randomised to one of two or more treatments, which are compared
Crossover: main alternative design to the parallel group designs where all participants receive the same treatment(s), but the order in which they receive them depends on the arm to which they are randomly assigned.
Superiority: trials designed to detect the superiority of a treatment compared to other treatment(s), in contrast to trials that are designed to assess non-inferiority or equivalence.

Blinding/masking (safeguard strategy against bias)

Open label: where researchers, providers, and participants are all aware of the treatment assigned because blinding to treatment assignment is difficult or impossible. Frequently applied in combination with blinded endpoint evaluation (see below).
Single-blind: where either the researcher collecting data, the healthcare provider, or the participant is blind to the treatment assigned.
Double-blind: where both the researcher collecting the data and the participant are blind to the treatment assigned.
Blinded endpoint: An independent investigator or preferably an endpoint committee who was blinded to the group assignment (and preferably time of assessment in designs involving a baseline assessment) did the assessment of clinical recovery in the study (blind end point assessment).

References

Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation*, *36*(3), 320–338. https://doi.org/10.1177/1098214014557009

Baumgaertner, A., Grewe, T., Ziegler, W., Floel, A., Springer, L., Martus, P., & Breitenstein, C. (2013). FCET2EC (From controlled experimental trial to= 2 everyday communication): How effective is intensive integrative therapy for stroke-induced chronic aphasia under routine clinical conditions? A study protocol for a randomized controlled trial. *Trials*, *14*(1), 1–8. https://doi.org/10.1186/1745-6215-14-308

Behn, N., Moss, B., McVicker, S., Roper, A., Northcott, S., Marshall, J., Thomas, S., Simpson, A., Flood, C., James, K., Goldsmith, K., & Hilari, K. (2021). SUpporting wellbeing through PEeR-befriending (SUPERB) feasibility trial: Fidelity of peer-befriending for people with aphasia. *BMJ Open*, *11*(8), e047994. https://doi.org/10.1136/bmjopen-2020-047994

Bellg, A., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., & Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology*, *23*(5), 443–451. https://doi.org/10.1037/0278-6133.23.5.443

Borelli, B., Sepinwall, D., Ernst, D., Bellg, A., Czajkowski, S., Breger, R., DeFrancesco, C., Levesque, C., Sharp, D. L., Ogedegbe, G., Resnick, B., & Orwig, D. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology*, *73*(5), 852–860. https://doi.org/10.1037/0022-006X.73.5.852

Borrelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry*, *71*(s1), S52–S63. http://dx.doi.org/10.1111/j.1752-7325.2011.00233.x

Breitenstein, C., Grewe, T., Flöel, A., Ziegler, W., Springer, L., Martus, P., Huber, W., Willmes, K., Ringelstein, E. B., Haeusler, K. G., Abel, S., Glindemann, R., Domahs, F., Regenbrecht, F., Schlenck, K.-J., Thomas, M., Obrig, H., de Langen, E., Rocker, R., … Bamborschke, S. (2017). Intensive speech and language therapy in patients with chronic aphasia after stroke: A randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *The Lancet*, *389*(10078), 1528–1538. https://doi.org/10.1016/S0140-6736(17)30067-3

Brogan, E., Ciccone, N., & Godecke, E. (2019). Treatment fidelity in aphasia randomised controlled trials. *Aphasiology*, *33*(7), 759–779. https://doi.org/10.1080/02687038.2019.1576442

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *BMJ*, *321*(7262), 694–696. https://doi.org/10.1136/bmj.321.7262.694

Carragher, M., Ryan, B., Worrall, L., Thomas, S., Rose, M., Simmons-Mackie, N., Khan, A., Hoffmann, T. C., Power, E., Togher, L., & Kneebone, I. (2019). Fidelity protocol for the action success knowledge (ASK) trial: A psychosocial intervention administered by speech and language therapists to prevent depression in people with post-stroke aphasia. *BMJ Open*, *9*(5), e023560. https://doi.org/10.1136/bmjopen-2018-023560

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*(1), 40. https://doi.org/10.1186/1748-5908-2-40

Cherney, L. R., Patterson, J. P., Raymer, A., Frymark, T., & Schooling, T. (2008). Evidence-based systematic review: Effects of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia. *Journal of Speech, Language, and Hearing Research*, *51*(5), 1282–1299. https://doi.org/10.1044/1092-4388(2008/07-0206)

Cherney, L. R., Simmons-Mackie, N., Raymer, A., Armstrong, E., & Holland, A. (2013). Systematic review of communication partner training in aphasia: Methodological quality. *International Journal of Speech-Language Pathology*, *15*(5), 535–545. https://doi.org/10.3109/17549507.2013.763289

Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., & Lancaster, G. A. (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *BMJ*, *355*. https://doi.org/10.1136/bmj.i523

Faroqi-Shah, Y., Frymark, T., Mullen, R., & Wang, B. (2010). Effect of treatment for bilingual individuals with aphasia: A systematic review of the evidence. *Journal of Neurolinguistics*, *23*(4), 319–341. https://doi.org/10.1016/j.jneuroling.2010.01.002

Feely, M., Seay, K. D., Lanier, P., Auslander, W., & Kohl, P. L. (2018). Measuring fidelity in research studies: A field guide to developing a comprehensive fidelity measurement system. *Child and Adolescent Social Work Journal*, *35*(2), 139–152. https://doi.org/10.1007/s10560-017-0512-6

Godecke, E., Armstrong, E., Rai, T., Ciccone, N., Rose, M. L., Middleton, S., Whitworth, A., Holland, A., Ellery, F., Hankey, G. J., Cadilhac, D. A., & Bernhardt, J. (2021). A randomized control trial of intensive aphasia therapy after acute stroke: The very early rehabilitation for speech (VERSE) study. *International Journal of Stroke*, *16*(5), 556–572. https://doi.org/10.1177/1747493020961926

Godecke, E., Armstrong, E., Rai, T., Middleton, S., Ciccone, N., Whitworth, A., Rose, M., Holland, A., Ellery, F., Hankey, G. J., Cadilhac, D. A., & Bernhardt, J. (2016). A randomized controlled trial of very early rehabilitation in speech after stroke. *International Journal of Stroke*, *11*(5), 586–592. https://doi.org/10.1177/1747493016641116

Godecke, E., Brogan, E., Ciccone, N., Rose, M. L., Armstrong, E., Whitworth, A., Ellery, F., Holland, A., Middleton, S., Rai, T., Hankey, G. J., Cadilhac, D., & Bernhardt, J. (under submission). *Treatment fidelity monitoring, reporting and findings in a complex aphasia intervention trial: A substudy of the very early rehabilitation in speech (VERSE) trial*.

Gwet, K. L. (2002). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods for Inter-rater Reliability Assessment*, *1*(6), 1–6. https://www.agreestat.com/papers/kappa_statistic_is_not_satisfactory.pdf

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (4th) Advanced Analytics, LLC.

Harrison, M. (2019). *Evaluating the intervention fidelity of self-managed computer therapy for aphasia post-stroke*. (Doctoral dissertation, University of Sheffield, Sheffield, United Kingdom).

Hart, T. (2009). Treatment definition in complex rehabilitation interventions. *Neuropsychological Rehabilitation*, *19*(6), 824–840. https://doi.org/10.1080/09602010902995945

Hilari, K., Behn, N., James, K., Northcott, S., Marshall, J., Thomas, S., Simpson, A., Moss, B., Flood, C., McVicker, S., & Goldsmith, K. (2021). Supporting wellbeing through peer-befriending (SUPERB) for people with aphasia: A feasibility randomised controlled trial. *Clinical Rehabilitation*, *35*(8), 1151–1163. https://doi.org/10.1177/0269215521995671

Hilari, K., Behn, N., Marshall, J., Simpson, A., Thomas, S., Northcott, S., Flood, C., McVicker, S., Jofre-Bonet, M., Moss, B., James, K., & Goldsmith, K. (2019). Adjustment with aphasia after stroke: Study protocol for a pilot feasibility randomised controlled trial for supporting wellbeing through PEeR befriending (SUPERB). *Pilot and Feasibility Studies*, *5*(1), 14. https://doi.org/10.1186/s40814-019-0397-6

Hildebrand, M. W., Host, H. H., Binder, E. F., Carpenter, B., Freedland, K. E., Morrow-Howell, N., Baum, C. M., Dore, P., & Lenze, E. J. (2012). Measuring treatment fidelity in a rehabilitation intervention study. *American Journal of Physical Medicine & Rehabilitation*, *91*(8), 715–724. https://doi.org/10.1097/PHM.0b013e31824ad462

Hinckley, J. J., & Douglas, N. F. (2013). Treatment fidelity: Its importance and reported frequency in aphasia treatment studies. *American Journal of Speech-Language Pathology*, *22*(2), S279–284. https://doi.org/10.1044/1058-0360(2012/12-0092)

Hoffman, T. C., Glasziou, P. P., Moher, D., Moher, D., Moher, D., Altman, D. G., Lamb, S. E., Lamb, S. E., Lamb, S. E., Lamb, S. E., Wyatt, J. C., Wyatt, J. C., Wyatt, J. C., Michie, S., Michie, S., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, *348*(mar07 3), g1687. https://doi.org/10.1136/bmj.g1687

Hohmann, A. A., & Shear, M. K. (2002). Community-based intervention research: Coping with the noise of real life in study design. *American Journal of Psychiatry*, *159*(2), 201–207. https://doi.org/10.1176/appi.ajp.159.2.201

Kaderavek, J. N., & Justice, L. M. (2010). Fidelity: An essential component of evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology*, *19*(4), 369–379. https://doi.org/10.1044/1058-0360(2010/09-0097)

Kazdin, A. E. (1986). Comparative outcome studies of psychotherapy: Methodological issues and strategies. *Journal of Consulting and Clinical Psychology*, *54*(1), 95–105. https://doi.org/10.1037/0022-006X.54.1.95

Lichstein, K. L., Riedel, B. W., & Grieve, R. (1994). Fair tests of clinical trials: A treatment implementation model. *Advances in Behavior Research and Therapy*, *16*(1), 1–29. https://doi.org/10.1016/0146-6402(94)90001-9

Loudon, K., Treweek, S., Sullivan, F., Donnan, P., Thorpe, K. E., & Zwarenstein, M. (2015). The PRECIS-2 tool: Designing trials that are fit for purpose. *BMJ*, *350*(may08 1), h2147. https://doi.org/10.1136/bmj.h2147

Markowitz, J. C., Spielman, L. A., Scarvalone, P. A., & Perry, S. W. (2000). Psychotherapy adherence of therapists treating HIV positive patients with depressive symptoms. *Journal of Psychotherapy Practice and Research*, *9*, 75–80. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3330590/

Marshall, J., Booth, T., Devane, N., Galliers, J., Greenwood, H., Hilari, K., Talbot, R., Wilson, S., Woolf, C., & Howell, P. (2016). Evaluating the benefits of aphasia intervention delivered in virtual reality: Results of a quasi-randomised study. *PLoS ONE*, *11*(8), e0160381. https://doi.org/10.1371/journal.pone.0160381

Masterson-Algar, P., Burton, C. R., Rycroft-Malone, J., Sackley, C. M., & Walker, M. F. (2014). Towards a programme theory for fidelity in the evaluation of complex interventions. *Journal of Evaluation in Clinical Practice*, *20*(4), 445–452. https://doi.org/10.1111/jep.12174

Miller, W. R., & Rollnick, S. (2014). The effectiveness and ineffectiveness of complex behavioral interventions: Impact of treatment fidelity. *Contemporary Clinical Trials*, *37*(2), 234–241. https://doi.org/10.1016/j.cct.2014.01.005

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*(3), 247–266. https://doi.org/10.1016/0272-7358(91)90103-2

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*(3), 315–340. https://doi.org/10.1177/109821400302400303

Orwin, R. G. (2000). Assessing program fidelity in substance abuse health services research. *Addiction*, *95*(Suppl. 3), S309–S327. https://doi.org/10.1080/09652140020004250

Palmer, R., Dimairo, M., Cooper, C., Enderby, P., Brady, M., Bowen, A., Latimer, N., Julious, S., Cross, E., Alshreef, A., Harrison, M., Bradley, E., Witts, H., & Chater, T. (2019). Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (big CACTUS): A multicentre, single-blinded, randomised controlled trial. *The Lancet Neurology*, *18*(9), 821–833. https://doi.org/10.1016/S1474-4422(19)30192-9

Palmer, R., Dimairo, E., Latimer, N., Cross, E., Brady, M., Enderby, P., Bowen, A., Julious, S., Harrison, M., Alshreef, A., Bradley, E., Bhadhuri, A., Chater, T., Hughes, H., Witts, H., Herbert, E., & Cooper, C. (2020). Computerised speech and language therapy or attention control added to usual care for people with long-term post-stroke aphasia: The big CACTUS three-arm RCT. *Health Technology Assessment*, *24*(19), 1–176. https://doi.org/10.3310/hta24190

Palmer, J., Parker, V., Barre, L., Mor, V., Volandes, A., Belanger, E., Loomer, L., McCreedy, E., & Mitchell, S. (2019). Understanding implementation fidelity in a pragmatic randomized clinical trial in the nursing home setting: A mixed-methods examination. *Trials*, *20*(1), 656. https://doi.org/10.1186/s13063-019-3725-5

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, *75*(6), 829–841. https://doi.org/10.1037/0022-006X.75.6.829

Purdy, M., Coppens, P., Madden, E. B., Mozeiko, J., Patterson, J., Wallace, S. E., & Freed, D. (2019). Reading comprehension treatment in aphasia: A systematic review. *Aphasiology*, *33*(6), 629–651. https://doi.org/10.1080/02687038.2018.1482405

Resnick, B., Inguito, P., Orwig, D., Yahiro, J. Y., Hawkes, W., Werner, M., Zimmerman, S., & Magaziner, J. (2005). Treatment fidelity in behavior change research: A case example. *Nursing Research*, *54*(2), 139–143. https://doi.org/10.1097/00006199-200503000-00010

Rose, M., Copland, D., Nickels, L., Togher, L., Meinzer, M., Rai, T., Cadilhac, D., Kim, J., Foster, A., Carragher, M., Hurley, M., & Godecke, E. (2019). Constraint-induced or multi-modal personalized aphasia rehabilitation (COMPARE): A randomized controlled trial for stroke-related chronic aphasia. *International Journal of Stroke*, *14*(9), 972–976. https://doi.org/10.1177/1747493019870401

Rose, M., Raymer, A. M., Lanyon, L. E., & Attard, M. C. (2013). A systematic review of gesture treatments for post-stroke aphasia. *Aphasiology*, *27*(9), 1090–1127. https://doi.org/10.1080/02687038.2013.805726

Rose, M., Togher, L., Godecke, E., Copland, D., Meinzer, M., Nickels, L., & Rai, T. (2019). *COMPARE fidelity monitoring plan version 7*. Accesssed at https://cloudstor.aarnet.edu.au/plus/s/k37xblJGkSsxLYR

Simmons-Mackie, N., Raymer, A., Armstrong, E., Holland, A., & Cherney, L. (2010). Communication partner training in aphasia: A systematic review. *Archives of Physical Medicine and Rehabilitation*, *91*(12), 1814–1837. https://doi.org/10.1016/j.apmr.2010.08.026

Simmons-Mackie, N., Raymer, A., & Cherney, L. R. (2016). Communication partner training in aphasia: An updated systematic review. *Archives of Physical Medicine and Rehabilitation*, *97*(12), 2202–2221 e2208. https://doi.org/10.1016/j.apmr.2016.03.023

Spell, L. A., Richardson, J. D., Basilakos, A., Stark, B. C., Teklehaimanot, A., Hillis, A. E., & Fridriksson, J. (2020). Developing, implementing, and improving assessment and treatment fidelity in clinical aphasia research. *American Journal of Speech-Language Pathology*, *29*(1), 286–298. https://doi.org/10.1044/2019_AJSLP-19-00126

Stirman, S. W., Baumann, A. A., Miller, C. J., Rabin, B., & Aarons, G. A. (2019). The FRAME: An expanded framework for reporting adaptations and modifications to evidence-based interventions. *Implementation Science*, *14*(1), 1–10. https://doi.org/10.1186/s13012-018-0842-6

Walker, M. F., Hoffmann, T. C., Brady, M. C., Dean, C. M., Eng, J. J., Farrin, A. J., Felix, C., Forster, A., Langhorne, P., Lynch, E. A., Radford, K. A., Sunnerhagen, K. S., & Watkins, C. L. (2017). Improving the development, monitoring and reporting of stroke rehabilitation research: Consensus-based core recommendations from the stroke recovery and rehabilitation roundtable. *ernational ournal ofInternational Journal of Stroke*, *12*(5), 472–479. https://doi.org/10.1177/1747493017711815

Walton, H., Spector, A., Williamson, M., Tombor, I., & Michie, S. (2020). Developing quality fidelity and engagement measures for complex health interventions. *British Journal of Health Psychology*, *25*(1), 39–60. https://doi.org/10.1111/bjhp.12394

Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology*, *13*(61), 1–7. https://doi.org/10.1186/1471-2288-13-61

Worrall, L., Ryan, B., Hudson, K., Kneebone, I., Simmons-Mackie, N., Khan, A., Hoffmann, T., Power, E., Togher, L., & Rose, M. (2016). Reducing the psychosocial impact of aphasia on mood and quality of life in people with aphasia and the impact of caregiving in family members through the aphasia action success knowledge (aphasia ASK) program: Study protocol for a randomized controlled trial. *Trials*, *17*(1), 153. https://doi.org/10.1186/s13063-016-1257-9

Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Haynes, B., Oxman, A. D., & Moher, D. (2008). Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *BMJ*, *337*(nov11 2), a2390. https://doi.org/10.1136/bmj.a2390