City, University of London Institutional Repository

A comparison of remote therapy, face to face therapy and an attention control intervention for

people with aphasia: A quasi-randomised controlled feasibility study

Celia Woolf[1]

Anna Caute[1]

Zula Haigh[1]

Julia Galliers[2]

Stephanie Wilson[2]

Awurabena Kessie[1]

Shashi Hirani[1]

Barbara Hegarty[3]

Jane Marshall[1]

[1] School of Health Sciences, City University London

[2] Centre for Human-Computer Interaction Design, City University London

[3] Homerton University Hospital, London

Corresponding author: Jane Marshall, Division of Language and Communication Science, City

University London, Northampton Square, London EC1V OHB. J.Marshall@city.ac.uk; 0207 040 4668

Abstract

Objective: To test the feasibility of a Randomised Controlled Trial comparing face to face and remotely delivered word finding therapy for people with aphasia

Design: A quasi-randomised controlled feasibility study comparing remote therapy delivered from a University lab, remote therapy delivered from a clinical site, face to face therapy and an attention control condition

Setting: A University lab and NHS outpatient service

Participants: Twenty-one people with aphasia following left hemisphere stroke

Interventions: Eight sessions of word finding therapy, delivered either face to face or remotely, were compared to an attention control condition comprising eight sessions of remotely delivered supported conversation. The remote conditions used mainstream video conferencing technology.

Outcome measures: Feasibility was assessed by recruitment and attrition rates, participant observations and interviews, and treatment fidelity checking. Effects of therapy on word retrieval were assessed by tests of picture naming and naming in conversation.

Results: Twenty-one participants were recruited over 17 months, with one lost at baseline. Compliance and satisfaction with the intervention was good. Treatment fidelity was high for both remote and face to face delivery (1251/1421 therapist behaviours were compliant with the protocol). Participants who received therapy improved on picture naming significantly more than controls (mean numerical gains: 20.2 (remote from University); 41 (remote from clinical site); 30.8 (face to face); 5.8 (attention control); p <.001). There were no significant differences between groups in the assessment of conversation.

Conclusions: Word finding therapy can be delivered via mainstream internet video conferencing. Therapy improved picture naming, but not naming in conversation.

Introduction

Telerehabilitation enables patients to access remote rehabilitation services in their own homes, typically by using internet video conferencing technologies. There are efficiency savings for both patients and service providers, mainly because the need to travel is eliminated. Such savings are particularly relevant in the context of stroke rehabilitation, where there are high levels of unmet need [1,2,3], and where demands on services are likely to increase. [4]

A number of studies have explored the use of telerehabilitation across a range of stroke services with promising findings. [5] However, the strength of evidence is low, with much of the data drawn from case series, [6,7] or feasibility studies. [8,9] A recent Cochrane review of telerehabilitation services for stroke concluded that there is insufficient evidence to guide practice. [10]

Applications of telerehabilitation in the domain of aphasia are even more preliminary. A number of studies have shown that remote language assessment is reliable and acceptable to participants. [11-15] There is also some evidence that remote administration of treatment can improve targeted skills, and achieve good levels of patient satisfaction. [16-23] Two studies compared face to face with remote delivery of aphasia therapy. [16, 18] Results showed that treatment gains were either no different across delivery modes [16] or marginally better in the remote condition. [18]

Although positive, the evidence base for telerehabilitation in aphasia is very limited. Most previous treatment studies involved case series designs, [17, 19, 20, 21] and the largest involved only eight participants. [18] The two studies that directly compared face to face with remote therapy employed cross over designs, in which participants received both forms of treatment. [16, 18] As a result, some of the gains may have been due to the cumulative, rather than independent effects of therapy (although Fridler et al [18] were careful to examine, and dismiss, treatment order effects).

Previous investigations of remote therapy have also typically employed bespoke technological platforms. The use of 'off the shelf' video conferencing systems, such as Skype[TM] and FaceTime®, has not been tested, despite the fact that these, unlike bespoke systems, are freely available. There is also a need to test these systems in routine clinical settings, as well as research laboratories.

The current study aimed to test the feasibility of a randomised controlled trial, comparing remote and face to face therapy. We additionally included an attention control condition, comprising remote supported conversation. The remote conditions employed a mainstream videoconferencing platform, and therapy was delivered from both a University lab and a clinical site. Therapy for word retrieval was employed, as word retrieval deficits are widespread in aphasia and respond well to intervention. [24, 25, 26] The treatment was specified in a manual and fidelity of delivery was examined.

The study addressed the following questions:

1. Can the same protocol of word finding therapy for people with aphasia be delivered face to face and remotely via mainstream internet videoconferencing technology?
2. Is treatment fidelity affected by delivery modes?
3. Do participants with aphasia comply with and express satisfaction with the treatment protocol; and do those in the remote condition find the technology accessible?
4. Does therapy improve word retrieval in naming tasks and conversation, and more so than an attention control intervention?

Method

The project received clearance from the Ethics Committee of the School of Health Sciences, City University London, the NRES Committee South East Coast – Surrey (12/LO/1932) and the Research and Development office of the clinical site used in the study.

This study was a quasi-randomised, repeated measures trial, comparing 4 intervention groups: (i) remote therapy delivered from a University lab (ii) remote therapy delivered from a clinical site, (iii) face to face therapy and (iv) an attention control condition. All participants were measured at four time points: week 1 (first baseline), week 4 (second baseline), week 8 (post therapy), week 14 (follow up). The study took place between March 2012 and November 2013.

Participants were recruited from community stroke groups in London, from the membership of a University aphasia research clinic, via an inner London NHS rehabilitation service and through self/relative referrals (e.g. from individuals who read about the project on line). The total recruitment period was 17 months. Participants were provided with information materials designed to be accessible to people with aphasia, [27] and given a verbal explanation of the project. They were given at least 24 hour consideration time before written consent was obtained. Participants without the capacity to consent were not approached or recruited.

The following eligibility criteria were identified at screening: participants were at least 6 months post a left hemisphere stroke; they had word finding difficulties due to aphasia (scoring between 20% and 70% correct on the spoken picture naming subtest of the Comprehensive Aphasia Test (CAT); [28] they demonstrated picture recognition and memory skills (scoring at least 70% on the CAT semantic and recognition memory subtests); they showed no signs of visual neglect (scoring within normal limits on the CAT line bisection test); they had no hearing loss greater than 40dB (established via pure tone audiometry); they had no secondary neurological diagnosis such as dementia; they were not receiving speech and language therapy elsewhere. Participants were also required to nominate a family member, friend or volunteer who could act as their partner in a conversation assessment and, if relevant, support their use of technology. Partners had no neurological impairment and no significant hearing loss.

Following screening, eligible participants were assigned to one of the 4 intervention groups. For logistical reasons[1] the first 5 recruits were assigned to remote therapy delivered from a University lab. The remaining recruits were randomly assigned to the other three interventions (see figure 1). Randomisation was conducted by a member of the team who was blind to referral and screening data using a computer generated randomisation table. Randomisation was blocked, to ensure that equal numbers were assigned to each condition. Consecutive recruits were referred to the team member who consulted the table to assign participants.

The study employed two remote therapy groups. Five participants were treated from a University lab, and five from a clinical site. This enabled us to explore the feasibility of using mainstream remote technologies from different settings. We envisaged that there might be additional barriers in the National Health Service context, e.g. relating to treatment space, technology availability or concerns about data protection.

Interventions

Those assigned to the therapy conditions received 8 one hour sessions of therapy delivered twice a week. Treatment aimed to improve spoken word production and followed a standard protocol, written in a manual. Each participant worked on 50 words, and each word was targeted at least once per session (the order of therapy stimuli varied systematically across sessions). Participants were provided with a work book, comprising pictures of their target words. The therapist worked with a corresponding book, which also delineated the tasks and cues that were to be used with each word. The therapy tasks were as follows:

Semantic verification: The therapist pointed to the target picture and asked two yes/no questions about the properties of the item (e.g. for *Lemon:* 'Can you squeeze it?'; 'Is it sweet?')

---

[1] This wing was funded by an award from the Bupa Foundation, received prior to the funding for the rest of the study

Picture naming: The therapist pointed to the picture and asked the participant to name it. If the participant was successful they were asked to repeat the word three times.

If the participant was unable to name the item the therapist offered the following cues (in the given order):

1. Semantic cue (e.g. for *Lemon*: 'We eat it with sugar on pancakes')

2. Sentence or phrase completion cue (e.g. 'Sour as a ....' *Lemon*)

3. First phoneme (sound) cue ('It begins with /l/')

4. First syllable cue ('It begins with /lɛ/')

5. Provision of the whole word for repetition.

Once the word was produced, the participant was asked to repeat it three times. If they were unable to say it, the therapist repeated the word three times.

All treatment sessions were video recorded for the purpose of fidelity checking.

Self administered practice: In addition to the therapist led treatment, participants in all therapy groups were provided with a simple, computer-based home practice task. It took the form of a Microsoft Powerpoint™ presentation that showed pictures of each target word. Participants were instructed to try naming each picture. If they were unable, they clicked on a blue microphone icon for a first phoneme cue. If this failed to elicit the word, they clicked on a green microphone icon for a whole word cue. Participants were encouraged to carry out the practice task on all their words at least 3 times a week.

Face to face therapy was delivered in a University lab. Participants travelled twice a week to the University for eight individual sessions. Remote therapy was administered via FaceTime using iPads. A prior consultation process, involving people with aphasia, identified this as the preferred

technology, on grounds of ease of access and quality of transmission. Prior to therapy, the remote participants, together with their partners, received at least one technology training session. This covered: Introduction to the iPad (e.g. turning on and off, using the stand, charging); connecting to the internet; receiving and ending FaceTime calls; adjusting the volume and trouble shooting (e.g. for when the connection is poor). Training was supported with simplified written and pictorial instructions, which were given to participants and their partners.

Following training, therapy was delivered twice a week for four weeks. The therapist called participants from the University or hospital clinic using a Mac Computer or iPad, and they received the calls at home using an iPad loaned from the study (one participant preferred to use Skype on his own PC). If necessary, participants were also provided with a mobile WiFi device and data allowance. Participants were instructed to use their workbook while receiving calls, so that the above naming and cueing protocol could be followed.

Participants assigned to the attention control condition received 8 remote conversation sessions. These were delivered over the internet using FaceTime. Sessions were scheduled twice a week (8 hours in total). As in the remote therapy groups, participants were loaned iPads and received initial training sessions in the use of the technology (content outlined above).

The conversations were conducted with students of speech and language therapy based in the University; i.e. students contacted participants from a Mac computer or iPad, and participants received the calls at home using the loaned iPad. Prior to their involvement, students received a half day training session in supported conversation and in the use of FaceTime. For example, this covered: how to initiate conversation topics and how to encourage initiation on the part of the aphasic person; how to adapt communication for people with aphasia; how to resolve communication breakdowns; how to use the technology, and how to support people with aphasia in its use. Students were provided with a handbook covering these points and offering further advice.

Fidelity checking

Fidelity checking of therapy was conducted by a researcher who was not otherwise involved in the study. A fidelity checklist was developed, which covered each stage of the treatment protocol as described in the manual, and recorded any deviations from it. Fifteen treatment sessions were checked (12.5% of the total). These comprised 5 from each therapy condition, with one video randomly selected from each participant. Three videos were double coded to check reliability, with 100% agreement. The second coder was not involved in treatment delivery.

Measures

Feasibility measures consisted of recruitment and attrition rates and the results of fidelity checking. Participants' compliance and satisfaction with the protocol was also assessed by observations and interviews that were conducted by a Human Computer Interaction (HCI) researcher who was not involved in treatment delivery. Fourteen participants were selected arbitrarily for observation and interview, with at least 3 from each condition. The selection was made by the HCI researcher, purely from their group and participant number. Observations and interviews took place towards the end of the intervention phase (in sessions 7 or 8). Participants in the remote conditions were also asked to rate aspects of the technology, such as the sound quality, and their level of competence in using it. Ratings were on a five point scale, supported by descriptive anchors. Compliance with the self-administered practice, for those in the therapy groups, was explored by asking participants to complete a simple diary of sessions undertaken. When these were not returned participants were simply asked about their homework.

The effect of therapy on word retrieval was assessed via two outcome measures that were administered twice before (week 1 and week 4) and twice after (week 8 and week 14) therapy, see figure 1. The primary measure was a test of spoken picture naming. This test is sensitive to therapy induced change [26] and the materials have at least 95% name agreement when tested with healthy

controls.[29] 190 pictures were presented for naming at week 1 and week 4. Each participant's

responses were examined in order to select 100 'difficult' items that would remain in the

experiment. These comprised: items that were failed in both tests, items that were failed in one test

and (if relevant) a random selection of correctly named items. These 100 words were tested again

after therapy (at week 8) and at follow up (week 14). For participants who received therapy, the 100

words were randomly assigned to two sub-sets. One formed the treated sub-set. The other was

untreated. T-test comparisons confirmed that the sub-sets did not differ with respect to word

frequency, familiarity, imageability or length.

In almost all cases, the naming assessments were administered by a member of the research team

who was not involved in the participant's therapy. However, it was difficult to blind assessors to the

time point. Therefore, assessments were video recorded and 28 of the week 8 and week 14

assessments (70% of the data) were double scored by a researcher who was blinded with respect to

the group allocation and time point. The percentage agreement across the blind and non blind

scorers was 91.8%; and the intraclass correlation coefficient was .945 (p<.001).

The secondary measure was an assessment of conversation. Participants were recorded in

conversation with their partner at each assessment point. The topic was unconstrained and the

researcher was not present. Conversations took place in participants' homes, or at the University or

Hospital clinic, and lasted approximately 10 minutes. The central 5 minutes was transcribed and

analysed using the POWERS procedure. [30] This procedure aims to quantify word retrieval during

conversation. The measure correlates with other assessments of word production [31], and is sensitive

to therapy induced change. [29]

Three POWERS indices were used to explore change over time:

Proportion of substantive turns: This was the proportion of conversational turns produced by the participant with aphasia that included at least one content word (a noun, verb, adjective, adverb or numeral).

Mean number of content words per turn: This was the total number of content words produced by the participant with aphasia divided by the number of their turns.

Mean number of nouns per turn: This was the total number of nouns produced by the participant with aphasia divided by the number of their turns.

The transcription and scoring was conducted by researchers who were blinded to group assignment and to the time point.

Results

Recruitment, Retention and Participant Characteristics

Figure 1 reports participant recruitment and retention through the study. Twenty one participants were recruited from a total of 45 screened, a conversion rate of 46.6%. Non recruitment was almost entirely due to participants not meeting the selection criteria. Attrition was low. One person did not complete the baseline tests for family reasons, and one failed to complete the week 14 conversation assessment due to hospitalisation. All participants who completed baseline received the allocated intervention.

Table 1 summarises the demographic and screening data for the 20 participants who received intervention. The table indicates that participants who received remote therapy from the University were, on average, older than the other groups and had more chronic aphasia. Their naming scores also appear lower. However, t-test comparisons revealed no significant differences between any of

the groups (p>.05) with respect to age, time post stroke onset, or any of the screening scores. When asked about prior computer experience, 71% of the participants indicated that they had used a computer before their stroke. Groups did not differ with respect to prior computer use.

Insert Figure 1 and Table 1 about here

Feasibility of the Intervention

Participants complied well with the protocol, with no missed sessions in any of the conditions. Human Computer Interaction observations showed that most of the participants in the remote groups mastered the technology and rated their competence highly, although two depended on their partner's support for some of the functions (see table 2). Most participants also gave good ratings for connectivity and for the visual and sound quality. Problems with connectivity were tolerated well, e.g. participants indicted that they were solved by redialling or moving the iPad to a different room. All participants in the therapy groups reported that they undertook at least 2 independent practice sessions per week.

Participant interviews asked about overall satisfaction with the intervention. Their response was supported by a 3 point scale, accompanied by descriptors and a happy, neutral or sad face. All bar one of the participants selected 'good' and the happy face. The exception (who received remote therapy from the University) indicated that her feelings were neutral.

Insert Table 2 about here

One feasibility issue emerged at the clinical site. Initially, permission to use a mainstream internet video conferencing platform was withheld. This, however, was overturned following discussion with a senior clinical manager (see discussion).

Treatment Fidelity

A total of 1421 therapist behaviours were coded during fidelity checking, across the 15 treatment videos. Of these 1251 (88%) were compliant with the protocol. The level of compliance was highest in remote therapy delivered from the University (93%). This compared to 82% in remote therapy from the clinical site and 86% in face to face therapy. Almost all deviations involved either the addition or omission of cues, driven by participant behaviours and preferences. For example, additions included the provision of further semantic cues or correction of articulation errors. Omissions mainly involved missing out a step in the cueing hierarchy.

Figure 2 reports the distribution of fidelity scores across individual sessions. The two low scoring sessions (55% and 57%) both involved remote therapy delivered from the clinical site. In one session the protocol deviation occurred because the patient became distressed. The other achieved low scores because phonological cues were omitted and naming was repeated twice rather than three times per item.

Impact of Therapy on Word Retrieval

Table 3 reports mean scores on the 100 item naming test, across the four groups. The first mixed factor ANOVA examined change over time in the total scores across the four intervention groups. Both main effects were significant (Time: $F_{(3,48)} = 134.4$, $p<0.001$, $\eta_p^2=.894$); Group: $F_{(3, 16)} = 11.62$, $p<0.001$, $\eta_p^2=.682$) and there was a significant interaction ($F_{(9, 48)} = 14.89$, $p<0.001$, $\eta_p^2=.736$). The results show that naming in all the therapy groups improved after intervention at week 8, with gains well maintained at week 14. In contrast, scores were largely unchanged in the attention control group. Table 4 illuminates these very contrastive gains. Follow-up single case analyses also showed that individual gains were highly significant for all participants in the therapy groups (Scores on treated words at week 4 and week 8 compared with McNemar chi square statistic, $p <0.005$).

13

Insert Table 3 and Table 4 about here

The second analysis examined the effect of treatment in more depth, so only included the three groups that received therapy. The within factors were time (4 levels: week 1, week 4, week 8 and week 14) and item (2 levels: treated words & untreated words). The between factor was group (3 levels: Remote therapy from the University; Remote therapy from the clinical site & Face to face therapy). This analysis produced a main effect of time ($F_{(3, 36)}$ = 129.37, $p<0.001$, η_p^2=.915). Pairwise comparisons showed no difference between week 1 and week 4 or between week 8 and week 14, but highly significant differences between week 4 and week 8 and between week 4 and week 14 ($p <0.001$). There was also a main effect of item ($F_{(1, 12)}$ = 57.887, $p<0.001$, Partial Eta Squared .828), indicating that treated words were named significantly better than untreated words. The main effect of group was also significant ($F_{(1, 12)}$ = 12.756, $p<0.001$, η_p^2= .68). Pairwise comparisons showed that the group receiving remote therapy from the clinical site scored more highly than both other groups ($p <0.05$), and that the face to face group scored more highly than the remote University group ($p<0.05$).

Three interactions were significant. The first was time by item ($F_{(3, 36)}$ = 36.04, $p<0.001$, η_p^2=.75), showing that treated items improved more over time than untreated items. The second was time by group ($F_{(6, 36)}$ = 5.51, $p<0.001$, η_p^2= .479). This showed that the groups differed with respect to their gains (see table 4). Those who received remote therapy from the clinical site improved the most. The third interaction was time by item by group ($F_{(6, 36)}$ = 2.62, $p<0.05$, η_p^2= .304). This showed that the groups differed in the degree to which gains occurred on treated or untreated items. Those who were treated remotely from the clinical site made gains in their naming of both treated and untreated words. While improvements in the other two groups were largely confined to treated words (see table 3).

14

Table 5 reports the mean scores on the three conversational indices across the four assessment time points. For the one participant who failed to complete week 14, the last available data point was entered into the analysis. Anova analyses were used to examine change over time on each index, with treatment mode as the between group variable. These analyses yielded no significant main effects and no interactions. Thus none of the intervention groups displayed improved word finding in conversation, as assessed by the POWERS procedure.

Insert Table 5 about here

Sample Size Calculation for a Non Inferiority Trial (Alpha 1%; Power 90%)

Power was calculated for a definitive non inferiority trial. This would aim to ensure that remote delivery of therapy does not result in worse outcomes than face to face delivery. The calculation was based purely on data from the primary outcome measure (picture naming). It assumed that a difference of 10% or more on this measure is clinically meaningful and assumed a SD of 18.3, which was the SD of words named correctly at week 14 by the 15 treated participants in the pilot. The calculation indicated that 176 participants (88 per group) are required to be 90% sure that the lower limit of a one-sided 99% confidence interval (or equivalently a 98% two-sided confidence interval) will be above the non-inferiority limit of -10.

Discussion

This study demonstrates that the delivery of remote word finding therapy for people with aphasia, using mainstream video conferencing technology, is feasible. Participants did not resist randomisation to the remote therapy condition. Indeed some expressed relief that they did not have to travel. All participants completed therapy, regardless of condition. Most of those who received remote intervention became independent in their use of the technology; and mean self

ratings for competence in using the technology and for the quality of transmission were high. Participants expressed good levels of satisfaction with all the interventions offered in this study, including the remote conditions.

There was no indication that remote delivery threatened treatment fidelity. Treatment fidelity is rarely assessed in aphasia therapy studies [32] and there is no consensus on how this should be done (although see suggestions in Kadaravek & Justice [33]). The current study coded individual therapist behaviours as compliant/not compliant with the treatment manual. This very stringent procedure showed that there were deviations from the protocol, e.g. because cues prescribed in the manual were omitted or augmented. Treating therapists indicated that this was in response to individual participant preferences. However, compliance was not adversely affected by remote delivery. Indeed, levels of compliance were highest in one of the remote conditions.

One feasibility problem that did emerge concerned the use of the mainstream internet technology from the clinical site. The hospital Information Technology department argued that their bespoke platform, delivered on personal computers, should be used, as this allowed all calls to be audited and excluded unauthorised users. The hospital branding would also reassure participants that calls were genuine. Unfortunately, the platform was not accessible to people with aphasia, largely because users had to respond to an email invitation in order to receive calls. The use of FaceTime was eventually approved by a senior clinical manager, because participants knew and recognised the therapist researchers and because the therapy sessions did not involve sensitive communication.

There are widespread concerns about data protection in the remote delivery of health services. For example, over half the respondents to a recent European Commission consultation into eHealth expressed anxieties about security and privacy when mobile devices are used in health care.[34] Clearly any follow up trial will need to tackle such concerns. It is encouraging that they could be

overcome in our feasibility study when the nature of the remote communication was clarified with managers.

Another feasibility issue relates to the recruitment rate, in that it took 17 months to recruit 21 participants, at a rate of 1.2 recruits per month.   The slow recruitment was partly due to the proportion of excluded participants, which exceeded 50%. The dominant reason for exclusion was a failure to satisfy the selection criteria, because word finding was either too severely or too minimally impaired. Only one exclusion occurred because the individual withheld consent.

Our recruitment and conversion rates were similar to those reported in other trials of speech and language therapy following stroke. [35, 36] For example, Palmer and colleagues [35] report a recruitment rate of 1.4 participants per site per month, with a site consisting of all speech and language therapy departments and voluntary sector organisations in a region; and in their trial 59% of those screened were excluded. It seems that recruitment to aphasia therapy trials is challenging. This may be because at least some people with aphasia are not in touch with referring agencies. The variability of aphasia also makes it very likely that a proportion of those referred to a trial will not meet the selection criteria.

Turning to clinical outcomes, both remote and face to face therapy benefited word retrieval, although only on our primary measure of picture naming. The groups who received therapy made average gains of between 20.2 and 41 on the 100 item naming test, compared to the mean gain of 5.8 achieved by the attention control group. The therapy gains were very well maintained at the final assessment, which took place 6 weeks after therapy had ceased. The lack of non-responders to therapy was a particularly impressive feature of the results, with all treated participants showing significant individual gains.

Although all the therapy groups improved, their gains varied. Those who were treated remotely from the clinical site improved the most, and demonstrated change on both treated and untreated words. In contrast, the group that received remote therapy from the University showed the smallest margin of change, with benefits that were more confined to treated words. These differences are difficult to interpret, given the low power of the study. They may reflect inequalities in baseline naming skills, [37] as the group that improved the most also achieved the highest naming scores in the pre therapy assessments. Crucially, the differences do not seem to be due to delivery mode. Overall, the ten people who received remote therapy achieved a very similar mean change to those who received face to face therapy.

The striking gains on the picture naming test were not matched by improvements on the secondary measure of conversation. None of the indices measured by the POWERS procedure showed change over time, and there were no group effects. Thus there was no evidence that any of the interventions employed in this study affected word finding in conversation.

Demonstrating therapy induced change in conversation is difficult, given the many variables that impact upon any test procedure. [29, 38] This null result may, therefore, reflect test sensitivity. However, the POWERS measure has been shown to correlate well with naming test scores, [31] and has demonstrated change in previous therapy studies. [29, 39, 40] An alternative explanation may lie in the therapy dose and/or content. Participants received only 8 hours of therapist led intervention, supplemented by homework practice. This low dose was clearly sufficient to improve picture naming, but may be insufficient to change conversation. The focus of therapy was also purely on single word production. The inclusion of more conversational tasks might have promoted greater generalisation. A third explanation may relate to the pattern of naming gains. With the exception of the group treated remotely from the clinical site, most participants only improved on the words that were practised in therapy. It seems likely that unconstrained conversation can only benefit when

naming gains extend to untreated words. Indeed the group treated from the clinical site was the only group that showed a trend towards improvement on the conversation measure (see table 5).

Before considering the implications of this study for any follow up, some substantial limitations need to be acknowledged. Assignment to groups was not fully random, since the first five recruits were allocated to remote therapy from the University. Although the intervention groups did not differ significantly on their demographic or screening scores, the power to detect change was low, given that there were only five people per group. More extensive testing might also have revealed disparities in baseline naming abilities or in other cognitive and linguistic factors that may have influenced the response to treatment.

Further limitations relate to blinding. Participants could not be blinded with respect to their intervention; i.e. they obviously knew whether they were receiving face to face or remote therapy. More positively, there was no indication that one intervention was perceived by participants as less valuable than the others. For example, participants did not resist randomisation to any group, and comments from at least some of the participants suggested that even the attention control was perceived as an active intervention. Although assessments were largely administered by non-treating therapists, assessors were not blind to the time point, with the concern that this might inflate post therapy scores. This problem was mitigated by blind scoring; i.e. 70% of the naming assessments conducted at weeks 8 and 14 were subjected to secondary scoring by a blinded assessor, and all conversation assessments were scored blind.

There were also limitations in the study sample. Participant numbers fell short of recommended sample sizes for pilot studies. [41,42] The age profile of the sample was also atypical of stroke. Twelve participants (60%) were below the age of 65, despite the fact that only a quarter of all strokes occur in this age group. [43] Most participants (71%) in our sample also had pre-stroke computer

experience. This probably reflected their age, given that this is a strong determinant of computer use. [44] A positive response to remote therapy might be more difficult to achieve with an older sample, and with individuals who have more limited prior exposure to computers, although it is worth noting that among our participants even those with little prior computer experience were able to master the technology.

The proportion of participants entering the study out of the total screened was low (46.6%). This gives rise to a further concern that therapy was tested with a highly selective sample. The majority of those who were excluded (21/24) did not meet the naming criteria. Relaxing these criteria would have permitted wider recruitment and possibly made the sample more reflective of the general population with aphasia. Conversely, as the therapy was focussed on word finding, it was important to ensure that this was a likely priority for those recruited.

In conclusion, we consider the implications for follow up research. This study found that word production, at least in picture naming tasks, was improved by therapy more than by the attention control condition, and this gain was regardless of delivery. However, the comparison between delivery modes needs further testing, via a fully powered non inferiority RCT. Our power calculation suggests that this should target 176 participants, with 88 per treatment arm. Achieving this number with our slow recruitment rate would be difficult. However, we only recruited from one NHS rehabilitation service, with all other referrals coming from the community. A larger trial would aim to recruit from all services in a region, as was the case in Palmer et al. [35] The target for the definitive trial could be met in 25 months, by recruiting from 5 regional sites at their rate of 1.4 recruits per month.

As already discussed, the sample tested in this feasibility study may not be representative of the wider stroke population, at least in terms of age. Further demographic variables, e.g. relating to

education and social and economic status were not collected, so there may have been additional imbalances. Recruiting from multiple sites, and from both health and community sources, should ensure that the sample in the larger trial is more representative.

The possible relationship between the therapy gain and baseline naming skills observed in this feasibility study suggests that naming should be extensively screened and randomisation stratified by naming severity. More extensive language and cognitive screening may also identify outcome predictors that could inform decisions about treatment candidacy. [37]

In terms of intervention, we have shown that remote therapy can be administered successfully not just from a University lab, but also from a routine clinical site. The latter can therefore be employed in the remote arm of the larger trial. Modification of the therapy content in the larger trial should be considered, to promote generalisation of gains to conversation. This might involve raising the dose and including communicative as well as single word tasks in therapy. [29]

Turning to assessment, a follow up study should explore word finding gains in conversation as well as naming. However, further constraints might be imposed on the conversational measure, e.g. so that the use of treated words can be explored. The assessment regime should also include a long term follow up, well beyond the 6 weeks used in the current study. Finally, a larger trial should include an economic assessment,[45] so that the cost effectiveness of face to face and remote therapy can be compared.

Clinical Messages

- Remote administration of word retrieval therapy for people with aphasia was feasible and acceptable to participants.

- Remote and face to face therapy improved word retrieval more than an attention control intervention.

- A definitive, non inferiority RCT trial is needed to compare remote and face to face therapy delivery; this will require 88 participants per treatment arm.

References

1. McKevitt C, Fudge N, Redfern J, Sheldenkar A, et al. Self-reported long-term needs after stroke. *Stroke* 2011; 42(5):1398-403.

2. Care Quality Commission. Supporting Life after Stoke National Report. A review of services for people who have had a stroke and their carers. http://www.cqc.org.uk/_db/_documents/Supporting_life_after_stroke_national_report.pdf (2011 accessed 23 August 2011)

3. Commission for Rural Communities. *Insights from stroke survivors and providers of stroke care in rural England: summary report.* 2010. Cheltenham: Commission for Rural Communities.

4. Feigin V, Lawes C, Bennett D, and Anderson C. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *The Lancet Neurology* 2003; 2: 43-53.

5. Johansson T and Wild C. (2011).Telerehabilitation in stroke care-a systematic review. *Journal of Telemedicine and Telecare* 2011; 17(1), 1-6. doi: 10.1258/jtt.2010.100105.

6. Holden M, Dyar T and Dayan-Cimadoro L. Telerehabilitation using a virtual environment improves upper extremity function in patients with stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 2007; 15: 36 – 42.

7. Piron L, Tonin P, Trivello E, Battisin L and Dam M. Motor tele-rehabilitation in post stroke patients. Medical informatics and the Internet in medicine 2004; 29: 119 – 125.

8. Huijgen B, Vollenbroek-Hutten M, Zampolini M. et al. Feasibility of a home based telerehabilitation system compared to usual care: arm/hand function in patients with stroke, traumatic brain injury and multiple sclerosis. *Journal of Telemedicine and Telecare* 2008; 14: 249 – 256.

9. Taylor D, Cameron J, Walsh L, et al. Exploring the feasibility of videoconference delivery of a self-management program to rural participants with stroke. *Telemedicine and eHealth* 2009: 15: 646 – 654.

10. Laver K, Schoene D, Crotty M, et al. Telerehabilitation services for stroke. *Cochrane Database of Systematic Reviews* 2013, Issue 12. Art. No.: CD010255. DOI: 10.1002/14651858.CD010255

11. Theodoros D, Hill A, Russell T, et al. Assessing acquired language disorders in adults via the Internet, *Journal of Telemedicine and e- Health* 2008; 14: 552-559.

12. Hill A, Theodoros D, Russell T. et al. The effects of aphasia severity on the ability to assess language disorders via telerehabilitation. *Aphasiology* 2009; 23: 627-642.

13. Palsbo S. Equivalence of functional communication assessment in speech pathology using videoconferencing. *Journal of Telemedicine and Telecare* 2007; 13: 40-3

14. Brennan D, Georgeadis A, Baron C and Barker L. The effect of videoconference-based telerehabilitation on story retelling performance by brain-injured subjects and its implications for remote speech-language therapy. *Telemedicine Journal and E-Health* 2004; 10: 147-154.

15. Georgeadis C, Brennan D, Barker M, and Baron C. Telerehabilitation and its effect on story retelling by adults with neurogenic communication disorders. *Aphasiology* 2004; 18: 639-652.

16. Agostini M, Garzon M, Benavides-Varela S, et al. Telerehabilitation in Poststroke Anomia. *Biomed Research International*, 2014(1)

17. Dechêne L, Tousignant M, Boissy P. et al. Simulated In-home Teletreatment for Anomia. *International Journal of Telerehabilitation* 2011; 3, 2, doi: 10.5195/ijt.2011.6075

18. Fridler N, Rosen K, Herzberg O, et al. Tele-Rehabilitation Therapy vs. Face-to-Face Therapy for Aphasic Patients. *eTELEMED 2012* : *The Fourth International Conference on eHealth, Telemedicine, and Social Medicine* 2012. ISBN: 978-1-61208-179-3

19. Furnas D and Edmonds L. The Effect of Computerized Verb Network Strengthening Treatment on Lexical Retrieval in Aphasia. *Aphasiology* 2014; 28: 401 – 420.

20. Goldberg S, Haley K and Jacks A. Script Training and Generalization for People with Aphasia. *American Journal of Speech-Language Pathology* 2012; 21: 222-238.

21. Mortley J, Wade J, Hughes A and Enderby P. Effectiveness of computerised rehabilitation for long term aphasia: A case series study. *British Journal of General Practice* 2004; 54: 856 – 857.

22. Mortley J, Wade J, and Enderby P. Superhighway to promoting a client-therapist partnership? Using the internet to deliver word-retrieval computer therapy. *Aphasiology* 2004; 18: 193 – 211.

23. Wade J, Mortley J. and Enderby P. Talk about IT: Views of people with aphasia and their partners on receiving remotely monitored computer-based word finding therapy. *Aphasiology* 2004; 17: 1031 – 1056.

24. Nickels L. Therapy for naming disorders: revisiting, revising and reviewing. *Aphasiology* 2002; 16: 935–980.

25. Doesborgh S,Van de Sandt-Koenderman M, Dippel D. et al. Effects of semantic treatment on verbal communication and linguistic processing in aphasia after stroke: a randomized controlled trial. *Stroke* 2004; 35: 141-6.

26. Best W, Greenwood A, Grassly J, Herbert R, Hickin J and Howard D, Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex* 2013; 49: 2345 – 2357.

27. Rose T, Worrall L, Hickson L and Hoffman T. Aphasia friendly written health information: Content and design characteristics. *International Journal of Speech-Language Pathology* 2011; 13: 335 – 347.

28. Swinburn K, Porter G, and Howard D. The Comprehensive Aphasia Test 2004; New York: Psychology Press.

29. Best W, Grassly J, Greenwood A, et al. A controlled study of changes in conversation following aphasia therapy for anomia. *Disability and Rehabilitation* 2011; 33(3): 229–242.

30. Herbert R, Best W, Hickin J et al. Profile of word errors and retrieval in speech (POWERS) 2012; JR Press.

31. Herbert R, Hickin J, Howard D, et al. Do picture naming tests provide a valid assessment of everyday functional lexical retrieval*? Aphasiology* 2008; 22: 184-203.

32. Hinckley J, and Douglas N. Treatment fidelity: Its importance and reported frequency in aphasia treatment studies. *American Journal of Speech-Language Pathology* 2013; 22: S279 – S284.

33. Kadaravek J. and Justice L. Fidelity: An essential component of evidence based practice in speech-language pathology. *American Journal of Speech-Language Pathology* 2010; 19: 369 – 379.

34. European Commission Summary Report on the Public Consultation on the Green Paper on Mobile Health, https://ec.europa.eu/digital-agenda/en/news/summary-report-public-consultation-green-paper-mobile-health (2015, accessed March 2015)

35. Palmer R, Enderby P, Cooper C, Latimer N, Julious S, Paterson G, Dimairo M, Dixon S, Mortley J, Hilton R, Delaney A and Hughes H, Computer therapy compared with usual care for people with long-standing aphasia poststroke: a pilot randomized controlled trial. *Stroke* 2012, 43: 1904-1911

36. Bowen A, Hesketh A, Patchick A, Young A, Davies L, Vail A, Long A, Watkins C, Wilkinson M, Pearl G, Lambon Ralph M, Tyrrell P. Effectiveness of enhanced communication therapy in the first four months after stroke for aphasia and dysarthria: a randomised controlled trial. *BMJ*, 2012, 345, e4407. DOI:10.1136/bmj.e4407

37. Lambon Ralph M, Snell C, Fillingham J et al. Predicting the outcome of anomia therapy for people with aphasia post CVA: Both language and cognitive status are key predictors. *Neuropsychological Rehabilitation* 2010; 20: 289 – 305.

38. Conroy P, Sage K, and Ralph M. Improved vocabulary production after naming therapy in aphasia: can gains in picture naming generalize to connected speech? *International Journal of Language and Communication Disorders* 2009; 44: 1036–1062.

39. Hickin J, Herbert R, Best W, et al. Efficacy of treatment: Effects on word retrieval and conversation. In Byng S, Swinburn K and Pound C (Eds) *Aphasia Therapy File (Vol. II).* Hove UK: Psychology Press 2007, pp 69 – 82.

40. Greenwood A, Grassly J, Hickin J, et al. Phonological and orthographic cueing therapy: a case of generalised improvement. *Aphasiology* 2010; 24 (9): 991-1016.

41. Julius S. Sample size of 12 per group as rule of thumb for a pilot study. *Pharmaceutical Statistics* 2005; 4: 287 – 291.

42. Billingham S, Whitehead A & Julious S An audit of sample sizes for pilot and feasibility trials being undertaken in the United Kingdom registered in the United Kingdom Clinical Research Network database. *BMC Medical Research Methodology* 2013; 13: 104.

43. The Stroke Association *Stroke Statistics*. 2012; The Stroke Association, UK.

44. Office for National Statistics Internet Access - Households and Individuals, http://www.ons.gov.uk/ons/rel/rdit2/internet-access---households-and-individuals/2013/stb-ia-2013.html (2013, accessed August 2014)

45. Medical Research Council A Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health. 2000.

Table 1: Participant Characteristics

	Remote from University	Remote from clinical site	Face to face	Attention control
Mean age (SD)	67.2 (6.98)	58.6 (14.38)	57.8 (15.14)	53 (13.93)
Mean months post stroke (SD)	53.4 (28.81)	31.8 (14.11)	35.2 (33.16)	20.2 (10.64)
Gender	3 Men 2 Women	4 Men 1 Woman	3 Men 2 Women	4 Men 1 Woman
Mean CAT semantic memory score (SD)	9 (1.22)	9.8 (.45)	8.4 (.89)	9 (1)
Mean CAT recognition memory score (SD)	9.4 (.89)	9 (.71)	9.6 (.55)	9.6 (.89)
Mean CAT naming score (SD)	15.4 (7.5)	27.4 (5.94)	20.2 (8.84)	15.6 (6.94)

Table 2: Summary of Human Computer Interaction observations and rating scores across 11 participants in the remote therapy and supported conversation conditions

Function	Observations	Mean Rating Score*(SD)
Starting the iPad	9 prompt and independent; 2 assisted by partners.	1.72 (1.42)
Starting FaceTime	8 prompt and independent; 1 achieved after errors; 2 assisted by partners	2.09 (1.7)
Answering the call	All prompt and independent	1.72 (.9)
Ending the call	Most commented that the therapist/students ended calls; 8 demonstrated awareness of how to end calls	2.23 (1.29)
Charging the iPad	Only five participants were observed on or asked about this function; All took responsibility for charging	Not rated
Connectivity	Connectivity was good in all observed sessions. 2 participants reported losses of connectivity in previous sessions which were solved by re-dialling or moving the iPad.	1.9 (1.37)
Sound quality	In all observed sessions sound was good; 7 recalled previous occasions when sound was poor but noted that this was the exception.	1.68 (.63)
Visual quality	Visual quality was good in all observed sessions; 6 reported that the screen occasionally froze but indicated that this resolved quickly.	1.45 (.69)

* 1 – 5 Rating Scale, with 1 positive

Table 3: Mean scores (SD) per group on the naming assessment over the four time points

	Week 1			Week 4			Week 8			Week 14		
	Treated N= 50	Untreated N = 50	Total N = 100	Treated N= 50	Untreated N = 50	Total N = 100	Treated N= 50	Untreated N = 50	Total N = 100	Treated N= 50	Untreated N = 50	Total N = 100
Remote from University	0.4 (.89)	1.0 (1.41)	1.4 (1.95)	.4 (.89)	.6 (1.34)	1.0 (2.24)	18.0 (4.95)	3.2 (3.70)	21.2 (5.89)	16.8 (7.63)	4.6 (2.7)	21.4 (7.23)
Remote from clinical site	5.2 (1.92)	5.8 (2.17)	11.0 (3.08)	8.6 (5.03)	6.4 (1.95)	15.0 (6.82)	31.8 (4.38)	24.2 (5.07)	56.0 (7.78)	30.0 (5.10)	25.2 (4.21)	55.2 (8.98)
Face to face	1.0 (2.24)	2.6 (3.44)	3.6 (5.50)	3.8 (4.66)	3.4 (3.91)	7.2 (8.41)	26.4 (9.45)	11.6 (7.89)	38.0 (16.4)	26.8 (11.78)	12.6 (6.8)	39.4 (17.88)
Attention control			4.8 (10.73)			3.8 (7.95)			9.6 (10.14)			9.8 (12.28)

N = Number of items that were treated and/or assessed

Table 4: Mean difference in total naming scores between week 4 and week 8, and week 4 and week 14 (95% CI)

	Week 4 vs Week 8	Week 4 vs Week 14
Remote therapy from University	20.2 (14.98 – 25.42)	20.4 (10.72 – 30.08)
Remote therapy from clinical site	41 (29.72 – 52.28)	40.2 (32.89 – 47.51)
Face to face therapy	30.8 (13.56 – 48.04)	32.2 (16.34 – 48.06)
Attention control	5.8 (.15 –11.44)	6 (-1.02 – 13.02)

Table 5: Results from the assessment of conversation: Mean scores (SD) per group for the POWERS indices over the four assessment time points

	Week 1			Week 4			Week 8			Week 14		
	Proportion substantive turns	Content words per turn	Nouns per turn	Proportion substantive turns	Content words per turn	Nouns per turn	Proportion substantive turns	Content words per turn	Nouns per turn	Proportion substantive turns	Content words per turn	Nouns per turn
Remote from University	.66 (.22)	2.92 (4.71)	1.34 (1.78)	.56 (.14)	4.24 (7.70)	2.03 (3.45)	.55 (.14)	3.09 (4.77)	1.65 (2.34)	.60 (.09)	3.65 (5.91)	1.96 (3.1)
Remote from clinical site	.62 (.18)	1.41 (.74)	.43 (.18)	.53 (.29)	1.56 (.96)	.5 (.47)	.86 (.09)	2.62 (1.81)	.70 (.41)	.74 (.15)	2.36 (1.80)	.63 (.22)
Face to face	.77 (.25)	2.39 (2.25)	1.16 (1.71)	.66 (.23)	1.98 (1.54)	.85 (.66)	.74 (.24)	1.9 (1.39)	.65 (.44)	.66 (.17)	1.99 (1.12)	.81 (.46)
Attention control	.68 (.22)	1.92 (.830	.73 (.76)	.66 (.21)	1.95 (.37)	.95 (.75)	.59 (.28)	2.49 (2.3)	.88 (.77)	.73 (.19)	2.16 (1.69)	.66 (.32)

Figure 1: Participant recruitment and retention
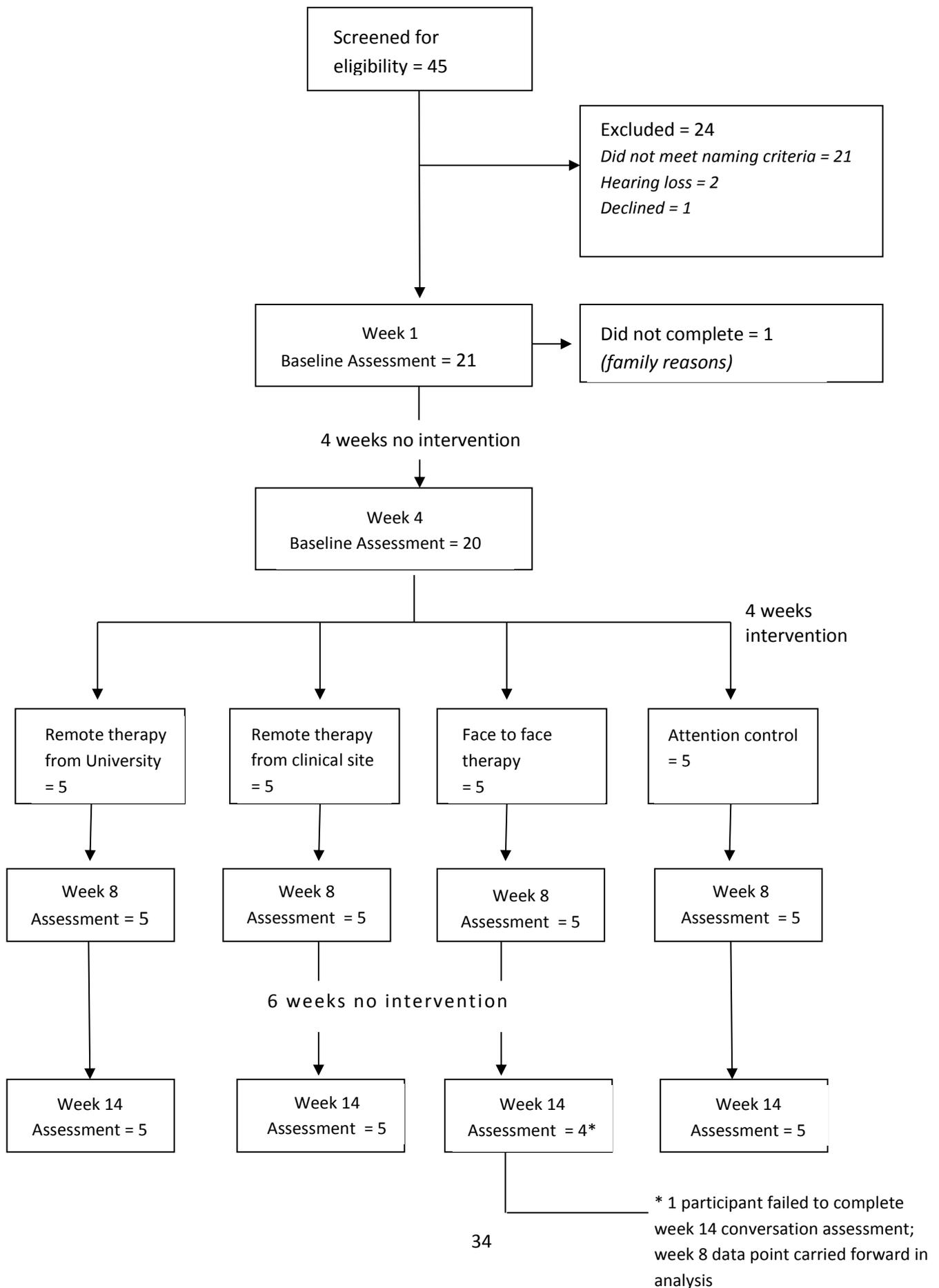
Figure 2: Distribution of fidelity scores (% of compliant therapist behaviours) across the 15 individual sessions