



This is a repository copy of *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.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/160216/>

Version: Published Version

Article:

Palmer, R. orcid.org/0000-0002-2335-7104, Dimairo, M. orcid.org/0000-0002-9311-6920, Latimer, N. orcid.org/0000-0001-5304-5585 et al. (14 more authors) (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). pp. 1-176. ISSN 1366-5278

<https://doi.org/10.3310/hta24190>

© Queen's Printer and Controller of HMSO 2020. This work was produced by Palmer et al. under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

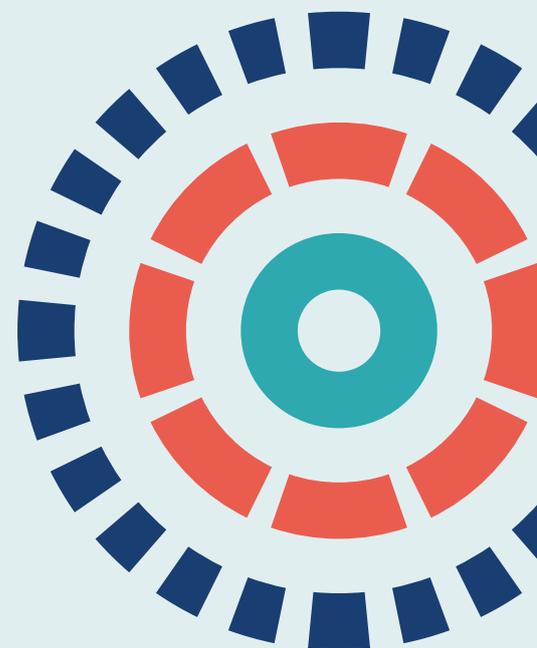
Health Technology Assessment

Volume 24 • Issue 19 • April 2020

ISSN 1366-5278

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

Rebecca Palmer, Munyaradzi Dimairo, Nicholas Latimer, Elizabeth Cross, Marian Brady, Pam Enderby, Audrey Bowen, Steven Julious, Madeleine Harrison, Abualbishr Alshreef, Ellen Bradley, Arjun Bhadhuri, Tim Chater, Helen Hughes, Helen Witts, Esther Herbert and Cindy Cooper



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

Rebecca Palmer^{1*} Munyaradzi Dimairo¹
Nicholas Latimer¹ Elizabeth Cross¹ Marian Brady²
Pam Enderby¹ Audrey Bowen³ Steven Julious¹
Madeleine Harrison¹ Abualbishr Alshreef¹
Ellen Bradley¹ Arjun Bhadhuri¹ Tim Chater¹
Helen Hughes^{1,4} Helen Witts^{1,5} Esther Herbert¹
and Cindy Cooper¹

¹School of Health and Related Research, University of Sheffield, Sheffield, UK

²Nursing, Midwifery and Allied Health Professions Research Unit, Glasgow Caledonian University, Glasgow, UK

³Division of Neuroscience & Experimental Psychology, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

⁴Speech and Language Therapy, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

⁵Speech and Language Therapy, Derbyshire Community Health Services NHS Foundation Trust, Chesterfield, UK

*Corresponding author

Declared competing interests of authors: All authors declare financial support from The Tavistock Trust for Aphasia. Rebecca Palmer was a National Institute for Health Research (NIHR)/Higher Education Funding Council for England-funded senior clinical academic lecturer until June 2017. She has current funding from the Stroke Association for a Doctor of Philosophy (PhD) student conducting work on fidelity to the intervention. The Stroke Association had previously funded early development work on the software used in the intervention but she was not involved in that. She was author of the intervention manual. Nicholas Latimer is supported by the NIHR (Post-doctoral Fellowship, reference PDF-2015-08-022) and is currently supported by Yorkshire Cancer Research (award S406NL). Pam Enderby has a patent on the Therapy Outcome Measures (2015) used in this trial from which she receives royalties. Audrey Bowen is funded by the Stroke Association and the NIHR Collaborations for Leadership in Applied Health Research and Care Greater Manchester. She co-authored the Communication Outcome after Stroke (COAST) and Carer COAST tools, which are patented. Madeleine Harrison receives PhD fellowship funding from the Stroke Association. Esther Herbert received a NIHR Research Methods Fellowship, outside the submitted work. Cindy Cooper sits on the NIHR Clinical Trials Unit (CTU) Standing Advisory Committee (2016 to present) and the UK Clinical Research Collaboration Registered CTU Network Executive Group (2015 to present).

Published April 2020

DOI: 10.3310/hta24190

This report should be referenced as follows:

Palmer R, Dimairo M, Latimer N, Cross E, Brady M, Enderby P, *et al.* 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 Technol Assess* 2020;**24**(19).

Health Technology Assessment is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.819

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 12/21/01. The contractual start date was in January 2014. The draft report began editorial review in July 2018 and was accepted for publication in October 2019. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

© Queen's Printer and Controller of HMSO 2020. This work was produced by Palmer *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Senior Clinical Researcher, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Director, NIHR Dissemination Centre, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Wellbeing Research, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

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

Rebecca Palmer ^{1*} Munyaradzi Dimairo ¹ Nicholas Latimer ¹
Elizabeth Cross ¹ Marian Brady ² Pam Enderby ¹ Audrey Bowen ³
Steven Julious ¹ Madeleine Harrison ¹ Abualbisher Alshreef ¹
Ellen Bradley ¹ Arjun Bhadhuri ¹ Tim Chater ¹ Helen Hughes ^{1,4}
Helen Witts ^{1,5} Esther Herbert ¹ and Cindy Cooper ¹

¹School of Health and Related Research, University of Sheffield, Sheffield, UK

²Nursing, Midwifery and Allied Health Professions Research Unit, Glasgow Caledonian University, Glasgow, UK

³Division of Neuroscience & Experimental Psychology, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

⁴Speech and Language Therapy, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

⁵Speech and Language Therapy, Derbyshire Community Health Services NHS Foundation Trust, Chesterfield, UK

*Corresponding author r.l.palmer@sheffield.ac.uk

Background: People with aphasia may improve their communication with speech and language therapy many months/years after stroke. However, NHS speech and language therapy reduces in availability over time post stroke.

Objective: This trial evaluated the clinical effectiveness and cost-effectiveness of self-managed computerised speech and language therapy to provide additional therapy.

Design: A pragmatic, superiority, single-blind, parallel-group, individually randomised (stratified block randomisation, stratified by word-finding severity and site) adjunct trial.

Setting: Twenty-one UK NHS speech and language therapy departments.

Participants: People with post-stroke aphasia (diagnosed by a speech and language therapist) with long-standing (> 4 months) word-finding difficulties.

Interventions: The groups were (1) usual care; (2) daily self-managed computerised word-finding therapy tailored by speech and language therapists and supported by volunteers/speech and language therapy assistants for 6 months plus usual care (computerised speech and language therapy); and (3) activity/attention control (completion of puzzles and receipt of telephone calls from a researcher for 6 months) plus usual care.

Main outcome measures: Co-primary outcomes – change in ability to find treated words of personal relevance in a bespoke naming test (impairment) and change in functional communication in conversation rated on the activity scale of the Therapy Outcome Measures (activity) 6 months after randomisation. A key secondary outcome was participant-rated perception of communication and quality of life using the Communication Outcomes After Stroke questionnaire at 6 months. Outcomes were assessed by speech

and language therapists using standardised procedures. Cost-effectiveness was estimated using treatment costs and an accessible EuroQol-5 Dimensions, five-level version, measuring quality-adjusted life-years.

Results: A total of 818 patients were assessed for eligibility and 278 participants were randomised between October 2014 and August 2016. A total of 240 participants (86 usual care, 83 computerised speech and language therapy, 71 attention control) contributed to modified intention-to-treat analysis at 6 months. The mean improvements in word-finding were 1.1% (standard deviation 11.2%) for usual care, 16.4% (standard deviation 15.3%) for computerised speech and language therapy and 2.4% (standard deviation 8.8%) for attention control. Computerised speech and language therapy improved word-finding 16.2% more than usual care did (95% confidence interval 12.7% to 19.6%; $p < 0.0001$) and 14.4% more than attention control did (95% confidence interval 10.8% to 18.1%). Most of this effect was maintained at 12 months ($n = 219$); the mean differences in change in word-finding score were 12.7% (95% confidence interval 8.7% to 16.7%) higher in the computerised speech and language therapy group ($n = 74$) than in the usual-care group ($n = 84$) and 9.3% (95% confidence interval 4.8% to 13.7%) higher in the computerised speech and language therapy group than in the attention control group ($n = 61$). Computerised speech and language therapy did not show significant improvements on the Therapy Outcome Measures or Communication Outcomes After Stroke scale compared with usual care or attention control. Primary cost-effectiveness analysis estimated an incremental cost per participant of £732.73 (95% credible interval £674.23 to £798.05). The incremental quality-adjusted life-year gain was 0.017 for computerised speech and language therapy compared with usual care, but its direction was uncertain (95% credible interval -0.05 to 0.10), resulting in an incremental cost-effectiveness ratio of £42,686 per quality-adjusted life-year gained. For mild and moderate word-finding difficulty subgroups, incremental cost-effectiveness ratios were £22,371 and £28,898 per quality-adjusted life-year gained, respectively, for computerised speech and language therapy compared with usual care.

Limitations: This trial excluded non-English-language speakers, the accessible EuroQol-5 Dimensions, five-level version, was not validated and the measurement of attention control fidelity was limited.

Conclusions: Computerised speech and language therapy enabled additional self-managed speech and language therapy, contributing to significant improvement in finding personally relevant words (as specifically targeted by computerised speech and language therapy) long term post stroke. Gains did not lead to improvements in conversation or quality of life. Cost-effectiveness is uncertain owing to uncertainty around the quality-adjusted life-year gain, but computerised speech and language therapy may be more cost-effective for participants with mild and moderate word-finding difficulties. Exploring ways of helping people with aphasia to use new words in functional communication contexts is a priority.

Trial registration: Current Controlled Trials ISRCTN68798818.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 24, No. 19. See the NIHR Journals Library website for further project information. The Tavistock Trust for Aphasia provided additional support to enable people in the control groups to experience the intervention after the trial had ended.

Contents

List of tables	xiii
List of figures	xvii
List of abbreviations	xix
Plain English summary	xxi
Scientific summary	xxiii
Chapter 1 Introduction	1
Post-stroke aphasia	1
Evidence for speech and language therapy	1
Delivery of therapy for persistent aphasia: the clinical problem	2
Potential solutions to increasing the amount of tailored therapy delivered for persistent aphasia without increasing demand on speech and language therapy resources	2
Intervention aimed at addressing delivery of long-term speech and language therapy for persistent aphasia investigated in the Big CACTUS trial: self-managed, computer aphasia therapy approach for persistent aphasia	2
Research rationale and objectives	3
Patient, carer and public involvement	4
Chapter 2 Methods	7
Trial design	7
Important changes to methods after trial commencement	7
Important protocol changes since version 1.0	7
Participants and eligibility criteria	8
Eligibility of providers	8
Settings and locations where the data were collected	9
<i>Participant identification</i>	9
<i>Screening for eligibility</i>	9
<i>Recruitment</i>	9
Interventions	10
<i>Usual-care control group</i>	10
<i>Self-managed computerised therapy intervention for word-finding (computerised speech and language therapy)</i>	13
<i>Attention control group</i>	17
Outcomes	17
<i>Primary outcome measures</i>	17
<i>Key secondary outcome measure</i>	18
<i>Secondary outcome measures</i>	19
<i>Staff training for delivery of outcome measures</i>	20
<i>Collection of demographic data</i>	20
<i>Recording of usual care</i>	20
Sample size	21
Interim analyses and stopping guidelines	21
<i>Numbers recruited</i>	22
<i>Recruitment as a percentage of the full-study recruitment targets</i>	22

CONTENTS

<i>Retention to first outcome measure time point at 6 months (primary outcome)</i>	22
<i>Identification and retention of volunteers</i>	22
<i>Summary</i>	22
Randomisation and concealment	22
Blinding	23
Chapter 3 Statistical analysis methods	25
Analysis populations	25
<i>Modified intention-to-treat set</i>	25
<i>Complete-case set</i>	25
<i>Per-protocol sets</i>	25
<i>Multiple imputation set</i>	26
<i>Linear interpolation set</i>	27
<i>Safety set</i>	27
Statistical considerations	27
<i>Approach to dealing with deaths after randomisation</i>	27
<i>Hochberg multiple testing procedure</i>	27
<i>Computation of summary outcome scores for analysis</i>	28
Statistical analysis methods	31
<i>Clinical effectiveness</i>	31
<i>Functional impact</i>	31
<i>Subgroup evaluation</i>	32
<i>Safety evaluation</i>	32
Chapter 4 Clinical effectiveness results	33
Introduction	33
Statistical results	33
<i>Screening and flow of trial participants</i>	33
<i>Characteristics of participating speech and language therapy departments in NHS trusts</i>	33
<i>Baseline demographics and characteristics of randomised participants</i>	33
<i>Reported cases of unblinding of outcome assessments</i>	41
<i>Usual-care speech and language therapy offered (fidelity/adherence to provision of usual care)</i>	41
<i>Fidelity to computerised speech and language therapy: adherence to practice and quality of intervention delivery</i>	42
<i>Per-protocol adherence to the attention control intervention</i>	44
<i>Response profiles of participants</i>	45
<i>Effect of the intervention on the co-primary and key secondary outcome measures</i>	47
<i>Impact of attrition and adherence on the co-primary and key secondary outcome measures</i>	48
<i>Subgroup influence on the effectiveness on the co-primary and key secondary outcome measures</i>	53
<i>Long-term effects of the intervention on the co-primary and key secondary outcomes</i>	55
<i>Effect of the intervention on generalisation to untreated words</i>	59
<i>Effect of the intervention on generalisation of treated words used in conversation</i>	61
<i>Carer-rated communication effectiveness and impact on carers' quality of life</i>	61
<i>Safety and negative effects</i>	63
Conclusions	69
Chapter 5 Health economics	73
Health economic analysis: summary of key points	73
Background	74
Overview of health economics methods	74

Methods: model-based analysis (primary analysis)	75
<i>Model design</i>	75
<i>Health-related quality of life</i>	76
<i>Resource use and costs</i>	78
<i>Addressing uncertainty</i>	79
<i>Subgroup analysis</i>	80
Methods: within-trial analysis (secondary analysis)	81
Results	81
<i>Model-based analysis</i>	81
<i>Within-trial analysis</i>	91
Chapter 6 Discussion and conclusions	95
Summary and interpretation of clinical findings	95
<i>Impact on word-finding</i>	95
<i>Functional communication, patient perception and quality of life</i>	95
<i>Generalisation to untreated words</i>	96
<i>Impact of aphasia profile, word-finding severity and length of time post stroke</i>	96
<i>Carer perception of communication effectiveness and impact on their quality of life</i>	96
<i>Safety</i>	96
Summary and interpretation of health economic findings	97
<i>Main findings</i>	97
<i>Explanation of findings</i>	98
<i>Subgroup analysis</i>	98
<i>Cost of computerised speech and language therapy intervention</i>	99
Fidelity to the interventions	99
<i>Quality of computerised speech and language therapy intervention delivery</i>	99
<i>Adherence to computerised speech and language therapy practice</i>	99
<i>Conclusion for computerised speech and language therapy</i>	99
<i>Fidelity to attention control</i>	99
<i>Fidelity to usual care</i>	100
Trial results in the context of other studies	100
<i>Age and gender</i>	100
<i>Computer use</i>	100
<i>Word-finding</i>	100
<i>Functional communication, participation and quality of life</i>	101
<i>Health economics</i>	101
Discussion of using co-primary outcomes	102
Strengths of the trial	102
<i>Strengths of the trial design</i>	103
<i>Patient and public involvement</i>	104
Limitations of the trial	105
Implications of the Big CACTUS trial findings	107
Further research	108
Acknowledgements	109
References	113
Appendix 1 List of participating NHS trusts	119
Appendix 2 Process for selection of Therapy Outcome Measures raters and scoring procedure	121

CONTENTS

Appendix 3 Definitions of adverse events and serious adverse events and categories of serious adverse event results	129
Appendix 4 Big CACTUS flow diagrams of activity	131
Appendix 5 Computer practice	137
Appendix 6 Unit costs	139
Appendix 7 EuroQol-5 Dimensions, five-level version completion information	141
Appendix 8 Secondary analysis model parameters: utility scores	143
Appendix 9 Computer and software costs scenario analysis	145
Appendix 10 Speech and language therapist, speech and language therapy assistant and volunteer costs	147
Appendix 11 Model parameters: resource use	149
Appendix 12 Model parameters used for base-case, secondary and subgroup analyses	151
Appendix 13 Expected value of information results	155
Appendix 14 Results of model-based secondary analyses	157
Appendix 15 Model-based subgroup analysis results	161
Appendix 16 Within-trial analysis: utility scores	165
Appendix 17 Within-trial analysis: histograms of cost distributions	169
Appendix 18 Within-trial analysis: base-case cost-effectiveness planes and cost-effectiveness acceptability curve	171
Appendix 19 Within-trial analysis: secondary analysis results	173
Appendix 20 Within-trial analysis: subgroup analysis results	175

List of tables

TABLE 1 Goals of therapy in UC	11
TABLE 2 Reasons for attrition	35
TABLE 3 Baseline demographic characteristics of randomised participants	35
TABLE 4 Baseline characteristics of randomised participants (outcome covariates)	38
TABLE 5 Cases of unblinding of outcome assessments	41
TABLE 6 Receivers of overall UC speech and language therapy	42
TABLE 7 Summary of the distribution of computer practice time ($n = 94$)	43
TABLE 8 Co-primary and key secondary outcomes at 6 months (mITT)	49
TABLE 9 Sensitivity analysis of the co-primary and key secondary outcomes at 6 months (mITT)	50
TABLE 10 Subgroup analysis results: co-primary and key secondary outcomes at 6 months	56
TABLE 11 Intervention effect on generalisation to untreated words at 6, 9 and 12 months	60
TABLE 12 Intervention effect on generalisation of treated words used in conversation	62
TABLE 13 Clinical improvement in generalisation of treated words used in conversation	63
TABLE 14 Intervention effect on the carer-rated communication effectiveness and carers' quality of life	64
TABLE 15 Negative effects of computer therapy (only those who used the computer)	65
TABLE 16 Negative effects of the computer therapy (treatment as randomised)	65
TABLE 17 Incidence of AEs (treatment as received)	66
TABLE 18 Incidence of AEs (treatment as randomised)	67
TABLE 19 Incidence of SAEs (treatment as received)	68
TABLE 20 Incidence of SAEs (treatment as randomised)	70
TABLE 21 Addressing uncertainty: secondary analyses	80
TABLE 22 Model parameters: transition probabilities	83
TABLE 23 Model parameters: utility scores (accessible EQ-5D-5L)	84

TABLE 24 Model-based results: base-case analysis	85
TABLE 25 Within-trial analysis results: base case	92
TABLE 26 Therapy outcome measures ratings of 10 videos by all raters: time point 1	122
TABLE 27 Inter-rater reliability of raters at time point 2 (excluding raters with low inter-rater reliability at time point 1 and/or low intrarater reliability)	123
TABLE 28 Intrarater reliability for each rater in <i>Tables 26</i> and <i>27</i> : TOMs benchmarking, reliability scores and intrarater reliability scores	124
TABLE 29 Category of SAE (treatment as received), unrelated or unlikely to be related to trial activity	130
TABLE 30 Unit costs applied for valuation of resource use	139
TABLE 31 Data completeness for EQ-5D-5L questionnaires	141
TABLE 32 Secondary analysis model parameters: utility scores (complete case, carer proxy, Hernandez-Alava <i>et al.</i> mapping and English EQ-5D-5L tariff)	143
TABLE 33 Computer and software costs scenario analyses	145
TABLE 34 The SLT, SLTA and volunteer costs	147
TABLE 35 Model parameters: resource use	149
TABLE 36 Model parameters used for base-case, secondary and subgroup analyses	152
TABLE 37 Expected value of partial perfect information: base-case analysis	156
TABLE 38 Cost-effectiveness results from base-case and secondary analyses: CSLT vs. UC model-based analysis	157
TABLE 39 Cost-effectiveness results from base-case and secondary analyses: CSLT vs. AC model-based analysis	158
TABLE 40 Cost-effectiveness results from base-case and secondary analyses: AC vs. UC model-based analysis	159
TABLE 41 Results of subgroup analyses comparing CSLT with UC: model-based analysis	162
TABLE 42 Results of subgroup analyses comparing CSLT with AC: model-based analysis	163
TABLE 43 Results of subgroup analyses comparing AC with UC: model-based analysis	164
TABLE 44 Mean and SD of accessible EQ-5D-5L scores by treatment group at each time point after missing EQ-5D-5L scores are imputed	165
TABLE 45 Mean, SD and sample size of accessible EQ-5D-5L scores by treatment group at each time point for complete cases	165

TABLE 46 Mean and SD of accessible EQ-5D-5L scores by treatment group at each time point for participants who provided a complete response at every data collection point	165
TABLE 47 Cost-effectiveness results from base-case and secondary within-trial analyses: CSLT vs. UC	173
TABLE 48 Cost-effectiveness results from base-case and secondary within-trial analyses: CSLT vs. AC	173
TABLE 49 Cost-effectiveness results from base-case and secondary within-trial analyses: AC vs. UC	174
TABLE 50 Results of subgroup analyses comparing CSLT with UC: within-trial analysis	175
TABLE 51 Results of subgroup analyses comparing CSLT with AC: within-trial analysis	175
TABLE 52 Results of subgroup analyses comparing AC with UC: within-trial analysis	176

List of figures

FIGURE 1 Interpretation of the Hochberg hierarchical sequential hypotheses testing strategy	28
FIGURE 2 Trial participant flow chart	34
FIGURE 3 Mean amounts of overall UC speech and language therapy (all participants)	42
FIGURE 4 Changes in word-finding over time stratified by the intervention	45
FIGURE 5 Changes in functional communication in conversation rated by TOMs over time, stratified by intervention	46
FIGURE 6 Changes in perception of social participation and quality of life rated by COAST over time, stratified by intervention	46
FIGURE 7 Mean word-finding ability over time stratified by the intervention	47
FIGURE 8 Mean functional communication in conversation over time stratified by the intervention	47
FIGURE 9 Mean COAST score over time, stratified by the intervention	51
FIGURE 10 Impact of attrition and intervention adherence on word-finding at 6 months	51
FIGURE 11 Impact of attrition and intervention adherence on functional communication at 6 months	52
FIGURE 12 Impact of attrition and intervention adherence on COAST at 6 months	52
FIGURE 13 Subgroup influence on word-finding results at 6 months	53
FIGURE 14 Subgroup influence on functional communication results at 6 months	54
FIGURE 15 Subgroup influence on COAST results at 6 months	55
FIGURE 16 Long-term intervention effect on word-finding of personal importance	58
FIGURE 17 Long-term intervention effect on functional communication	58
FIGURE 18 Long-term intervention effect on the COAST	59
FIGURE 19 Mean response profile in word-finding of treated words over time	61
FIGURE 20 Markov model structure	76
FIGURE 21 Cost-effectiveness plane: base-case analysis – CSLT vs. UC	86
FIGURE 22 Cost-effectiveness plane: base-case analysis – CSLT vs. AC	86

FIGURE 23 Cost-effectiveness plane: base-case analysis – AC vs. UC	87
FIGURE 24 Cost-effectiveness acceptability curve: base-case analysis	87
FIGURE 25 Means and 95% CIs of accessible EQ-5D-5L scores by treatment group at each time point after missing EQ-5D-5L scores are imputed	91
FIGURE 26 Computer therapy practice time per participant over time	137
FIGURE 27 Overall EVPI (per patient)	156
FIGURE 28 Means and 95% CIs of accessible EQ-5D-5L scores by treatment group at each time point for complete cases	166
FIGURE 29 Means and 95% CIs of accessible EQ-5D-5L scores by treatment group at each time point for participants who provided a complete response at every data collection point	166
FIGURE 30 Means and 95% CIs of proxy EQ-5D-5L scores by treatment group at each time point (complete cases)	166
FIGURE 31 Means and 95% CIs of carer EQ-5D-5L scores by treatment group at each time point (complete cases)	167
FIGURE 32 Histogram of CSLT costs	169
FIGURE 33 Histogram of AC costs	169
FIGURE 34 Cost-effectiveness plane: within-trial analysis, base case (CSLT vs. UC)	171
FIGURE 35 Cost-effectiveness plane: within-trial analysis, base case (CSLT vs. AC)	171
FIGURE 36 Cost-effectiveness plane: within-trial analysis, base case (AC vs. UC)	172
FIGURE 37 Cost-effectiveness acceptability curve: within-trial analysis, base case	172

List of abbreviations

AC	attention control	EVPI	expected value of perfect information
ACT NoW	Assessing Communication Therapy in the North West	HRQoL	health-related quality of life
AE	adverse event	ICER	incremental cost-effectiveness ratio
app	application	ICF	<i>International Classification of Functioning, Disability and Health</i>
CaCOAST	Carer Communication Outcome after Stroke	IQR	interquartile range
CaCOAST 15	First 15 items of the Carer Communication Outcome after Stroke	IR	incidence rate
CaCOAST 5	Last five items of the Carer Communication Outcome after Stroke	IRR	incidence rate ratio
CACTUS	Clinical and cost-effectiveness of aphasia computer treatment versus usual stimulation or attention control long term post-stroke	ITT	intention to treat
CarerQoL	Care-related Quality of Life instrument	MDC	mean difference in change
CAT	Comprehensive Aphasia Test	MI	multiple imputation
CEAC	cost-effectiveness acceptability curve	mITT	modified intention to treat
CI	confidence interval	NICE	National Institute for Health and Care Excellence
COAST	Communication Outcome after Stroke	OR	odds ratio
CONSORT	Consolidated Standards of Reporting Trials	PC	personal computer
CrI	credible interval	PCPI	patient, carer and public involvement
CSLT	computerised speech and language therapy	PhD	Doctor of Philosophy
CTRU	Clinical Trials Research Unit	PPI	patient and public involvement
DMEC	Data Monitoring and Ethics Committee	PSS	Personal Social Services
EQ-5D-3L	EuroQol-5 Dimensions, three-level version	PSSRU	Personal Social Services Research Unit
EQ-5D-5L	EuroQol-5 Dimensions, five-level version	QALY	quality-adjusted life-year
		RCT	randomised controlled trial
		RELEASE	Rehabilitation and recovery of people with aphasia after stroke
		SAE	serious adverse event
		SAP	statistical analysis plan
		SD	standard deviation
		SLT	speech and language therapist
		SLTA	speech and language therapy assistant

LIST OF ABBREVIATIONS

SUR	seemingly unrelated regression	TSC	Trial Steering Committee
TIDieR	Template for Intervention Description and Replication	UC	usual care
TOM	Therapy Outcome Measure	VAS	visual analogue scale

Plain English summary

Aphasia is a communication disorder that can be caused by a stroke. It affects a person's understanding of spoken words and their talking, reading and writing abilities. Communication may improve months, or years, after a stroke with speech and language therapy. Many patients want more speech and language therapy than the NHS can provide.

The Big CACTUS (clinical and cost-effectiveness of aphasia computer treatment versus usual stimulation or attention control long term post-stroke) trial evaluated the use of speech and language therapy software for people with aphasia to practise finding words independently at home on their own computer or one loaned by the NHS.

People with aphasia who had had a stroke at least 4 months previously were randomly allocated to one of three groups:

1. usual speech and language therapy care
2. daily use of computerised speech and language therapy for 6 months, tailored by a speech and language therapist and supported by a volunteer or speech and language therapy assistant
3. daily completion of puzzles and supportive telephone calls from a researcher to mimic the activity/attention the computerised speech and language therapy group received.

All groups received usual speech and language therapy.

A total of 278 people with aphasia took part in this trial, from 21 UK NHS speech and language therapy departments. They had their strokes between 4 months and 36 years previously. Computerised speech and language therapy enabled more practice (28 hours on average) than usual speech and language therapy (3.8 hours). The computerised speech and language therapy group significantly improved their ability to say words they chose to practise compared with those in the usual speech and language therapy or puzzle book groups.

Although computerised speech and language therapy can help people with aphasia to learn new words for years after stroke, no improvements in conversation or quality of life were seen. The cost-effectiveness for the NHS is still uncertain. However, our best estimate is that it is unlikely to be cost-effective for everyone with aphasia, but it may be cost-effective for people with mild and moderate word-finding difficulties. Next steps will focus on how to encourage use of new words in conversation to have an impact on quality of life.

Scientific summary

Background

More than one-third of stroke survivors acquire aphasia. This language disorder affects a person's ability to understand spoken language, talk, read and write, leading to frustration, isolation, low mood and difficulty carrying out roles at work, in the family and in the wider community. People with aphasia often want more speech and language therapy than they have access to from NHS services in the long term post stroke despite evidence that improvements can continue for years with speech and language therapy.

Self-managed use of specialist aphasia computer software provides a potentially low-cost option to enable more language practice without increasing demands on limited speech and language therapy resources. Computer software can provide the opportunity for repetitive practice of meaningful language material and feedback on success to the person with aphasia (adhering to principles of experience-dependent neuroplasticity) in their homes without a speech and language therapist present. The pilot study suggested feasibility, acceptability and potential clinical effectiveness and cost-effectiveness of an approach to self-managed computerised word-finding therapy. Big CACTUS (Clinical and cost-effectiveness of aphasia computer treatment versus usual stimulation or attention control long term post-stroke) was the first trial to evaluate the clinical effectiveness and cost-effectiveness of such an approach to providing long-term aphasia therapy in a full randomised controlled trial.

Objectives

The aim was to provide definitive evidence of whether or not self-managed computerised speech and language therapy for word-finding for persisting post-stroke aphasia in addition to usual care was clinically effective and cost-effective, when compared with usual care alone or attention control. The main research objectives were to establish whether or not:

1. self-managed computerised speech and language therapy for word-finding (computerised speech and language therapy) increases the ability of people with aphasia to retrieve vocabulary of personal importance
2. computerised speech and language therapy improves functional communication ability in conversation
3. patients receiving computerised speech and language therapy perceive greater changes in social participation in daily activities and quality of life
4. computerised speech and language therapy is cost-effective for persistent aphasia post stroke
5. any effects of computerised speech and language therapy are evident 12 months after therapy has begun.

This trial also sought to investigate:

6. functional use of treated words in conversation
7. generalisation to retrieval of untreated words
8. carer perception of communication effectiveness and the impact on carers' quality of life
9. negative effects of computerised speech and language therapy.

Methods

Design

Big CACTUS was a pragmatic, superiority, individually randomised, single-blind (blinded outcome assessors), parallel-group randomised controlled adjunct trial.

Setting

The trial was conducted in 21 NHS speech and language therapy departments across the UK. Participants were treated by speech and language therapists working in routine clinical practice.

Participants (eligibility and recruitment)

Participants were eligible if they were aged > 18 years and had aphasia subsequent to stroke(s) at least 4 months prior to randomisation. Participants were required to have word-finding difficulties (demonstrated by a score of 5–43 out of 48 on the Naming Objects test of the Comprehensive Aphasia Test) and to have sufficient cognitive and visual ability to use the computer software. They were excluded if they had a premorbid speech and language disorder caused by a neurological deficit other than stroke, required treatment in a language other than English or were currently using a computer program aimed at improving word-finding.

Eligible participants were identified from current and past speech and language therapy caseloads and from voluntary sector support groups. Speech and language therapists used a consent support tool to identify the support required for each individual to provide informed consent. This also identified those who were unable to provide consent with support and required a carer declaration of belief that they wished to take part (or a relative to provide consent in Scotland).

Interventions

Participants were randomised to one of three groups. All groups received usual care.

Usual care

This was the usual speech and language therapy being provided to each individual (for some, this entailed no speech and language therapy input). Therapy was provided by qualified speech and language therapists or speech and language therapy assistants face to face in either one-to-one or group sessions.

Computerised speech and language therapy

The computerised speech and language therapy intervention focused on the treatment of word-finding, a difficulty experienced by most people with aphasia. The approach evaluated had four key components:

1. Specialist aphasia software – StepByStep© version 5.0 (Steps Consulting Ltd, Acton Turville, UK) was chosen as it allowed for tailoring and selection of personally relevant words, and provided feedback on whether or not words produced were correct, adhering to principles of experience-dependent neuroplasticity.
2. A speech and language therapist assessed the participant and selected and tailored the software exercises and word-finding cues according to the language profile on the baseline language assessment. The speech and language therapist supported the participant to select 100 words of personal relevance for practice and trained and supported a volunteer/speech and language therapy assistant to provide support to the participant.
3. Self-managed, independent practice of the therapy exercises at home by the participant was recommended for 20–30 minutes per day over a 6-month period.
4. A volunteer/speech and language therapy assistant provided monthly support including encouraging practice, moving on to harder exercises, practice using the new words in functional contexts and adding new words if needed.

The speech and language therapists attended 1 day of training on how to use and tailor the software. They were given an intervention manual.

Attention control

This group aimed to control for the additional activity and attention received in the computerised speech and language therapy group. Puzzle books were provided each month according to level of difficulty needed and participant interests. Participants were encouraged to complete one puzzle per day for the 6-month period. A member of the research team telephoned the participant for a supportive chat once a month.

Randomisation

Participants were randomised to one of the three groups using a fixed 1 : 1 : 1 allocation ratio by a centralised web-based randomisation system. The randomisation schedule was generated using stratified block randomisation stratified by centre and severity of word-finding difficulty at baseline.

Outcomes

The trial had co-primary outcomes looking for change in both impairment and activity. Impairment was defined as the ability to use vocabulary of personal importance, measured by naming pictures of 100 words chosen for use in therapy. Activity was defined as functional communication ability in conversation measured by 10-minute videos of conversations structured around topics of personal importance, rated using the activity scale of the Therapy Outcome Measures. A key secondary outcome was change in participants' own perception of communication-related social participation and quality of life using the Communication Outcomes After Stroke patient-reported outcome measure. All outcome measures were undertaken at baseline prior to randomisation, 6 months post randomisation (end of treatment – primary time point) and at 9 and 12 months post randomisation. Further secondary outcomes included health-related quality of life and resource use to estimate the cost-effectiveness of computerised speech and language therapy compared with usual care and activity/attention control. An unofficial accessible variant of the EuroQol-5 Dimensions, five-level version, was developed for this trial so that participants were able to rate their own quality of life for quality-adjusted life-year estimation. The standard EuroQol-5 Dimensions, five-level version, was also completed by proxy by carers.

Additional outcomes included the use of treated words in the videoed conversations; generalisation of the therapy to untreated words using the Object Naming test of the Comprehensive Aphasia Test; and carer perception of communication effectiveness and impact on their own quality of life using the Carer Communication Outcomes After Stroke. Negative effects of the computerised speech and language therapy were recorded using a questionnaire for computerised speech and language therapy participants, and through collection of adverse events and serious adverse events for all groups.

Blinding

The trial was single blind as participants knew their allocated treatment group. However, all outcome measures at all time points were conducted by speech and language therapy assessors who were blind to the treatment group.

Sample size

The trial aimed to recruit 285 participants (95 per group) to address both co-primary objectives with 90% power for a 5% two-sided test adjusted for a 15% drop-out rate. A 10% mean difference in change in word-finding and a Therapy Outcome Measures effect size of 0.45 were assumed to be minimal clinically important differences to detect. This sample size had 83% power to address the key secondary objective on change in Communication Outcomes After Stroke for a 5% two-sided test assuming a 7.2% mean difference in change as clinically worthwhile.

Analysis

The analysis of the primary outcome measures was based on a modified intention-to-treat principle detailed in the full report. A multiple linear regression model adjusted for stratification factors was used. A sensitivity analysis was conducted exploring the impact of missing data and heterogeneity of treatment effect across predefined subgroups (word-finding severity, time post stroke and comprehension ability).

The primary health economic analysis was a model-based cost-utility analysis adopting a lifetime time horizon and an NHS payer perspective. Cost-effectiveness is expressed in terms of the incremental cost-effectiveness ratio, that is cost per quality-adjusted life-year gained. Secondary analyses included within-trial analysis, a broader perspective (including volunteer costs) and analysis of subgroups.

Results

A total of 995 potential participants were screened, of whom 288 (29%) consented and 278 (28%) were randomised, slightly lower than the target of 285 but sufficient to address the co-primary and key secondary research questions with intended statistical power owing to the drop-out rate (9%) being lower than anticipated (15%). A total of 240 participants were included in the modified intention-to-treat analysis: usual care, $n = 86$; attention control, $n = 71$; and computerised speech and language therapy, $n = 83$.

The mean age of participants was 65.4 years (range 23–92 years) and 61% of participants were male. Forty-four per cent of participants had mild word-finding difficulties (Comprehensive Aphasia Test Naming Objects score of 31–43), 30% had moderate difficulties (score of 18–30) and 26% had severe difficulties (score of 5–17). Participants were 2 years (median) post stroke (range 4 months to 36 years). The characteristics of groups were broadly similar at baseline.

Computerised speech and language therapy participants practised computer exercises for 28 hours (mean). The quality of delivery of computerised speech and language therapy in terms of tailoring the software and provision of support to the participants was good. However, 85% of participants were assisted to practise using their new words in functional contexts for only 45 minutes (median) in total over 6 months. Similar mean amounts of usual care were received by all groups across the 6-month intervention period (computerised speech and language therapy, 3.2 hours; usual care, 3.8 hours; and attention control, 3.2 hours).

The mean improvement in word-finding at 6 months was 1.1% (standard deviation 11.2%), 2.4% (standard deviation 8.8%) and 16.4% (standard deviation 15.3%) based on 86, 71 and 83 participants in the usual care, attention control and computerised speech and language therapy groups, respectively. On average, computerised speech and language therapy improved word-finding by 16.2% more than usual care (95% confidence interval 12.7% to 19.6%; $p < 0.0001$) and by 14.4% more than attention control (95% confidence interval 10.8% to 18.1%), indicating that the effect was not attributable only to activity/attention. Most of this effect was maintained at 9 and 12 months. Computerised speech and language therapy improvement in word-finding was broadly consistent regardless of time post stroke.

However, improvements in functional communication were negligible and very similar across groups. The mean difference in change between the computerised speech and language therapy group and the usual-care group was -0.03 (95% confidence interval -0.21 to 0.14 ; $p = 0.709$) and between the computerised speech and language therapy group and attention control group was -0.01 (95% confidence interval -0.20 to 0.18). Similarly, there was insufficient evidence to suggest that computerised speech and language therapy improved participants' perceptions of their communication ability and participation or its impact on their life. The computerised speech and language therapy group mean improvement in Communication Outcomes After Stroke was only 0.5% (95% confidence interval -3.1% to 4.1%) and 3.8% (95% confidence interval -0.0% to 7.5%) compared with usual care and attention control, respectively.

On average, there was no improvement in treated words used in conversation across groups. However, close to 1 out of 10 participants in the usual-care and attention control groups used at least five more treated words in conversation at 6 months than at baseline, compared with approximately 3 out of 10 participants in the computerised speech and language therapy group. In addition, there was insufficient evidence to suggest that improved word-finding of treated words generalises to untreated words.

Subgroup analyses indicated that the effect of computerised speech and language therapy on word-finding was slightly higher for participants with mild word-finding difficulties and for those with verbal comprehension within normal limits.

Small differences were seen in carers' perception of communication effectiveness: 4.6% in favour of computerised speech and language therapy compared with usual care (95% confidence interval 0.3% to 9.0%) at 6 months, and 5.1% in favour of computerised speech and language therapy compared with attention control (95% confidence interval 0.5% to 9.7%). The differences were not maintained at 9 and 12 months. Improvement in carers' quality of life in the computerised speech and language therapy group compared with the usual-care group was 5.3% (95% confidence interval -1.1% to 11.7%), but only 0.3% (95% confidence interval -6.4% to 6.9%) compared with the attention control group at 6 months.

Negative effects of computerised speech and language therapy were low: 27% of participants in the computerised speech and language therapy group reported fatigue or anxiety at some point, which translates to one event per person per year. Differences in the incidences of adverse events and serious adverse events were similar between groups.

The computerised speech and language therapy was a low-cost intervention, at £733 per person. The cost of delivering the same average amount (28 hours) of therapy face to face would be £1400. The primary cost-effectiveness analysis estimated an incremental cost per participant of £732.73 (95% credible interval £674.23 to £798.05) and an incremental quality-adjusted life-year gain of 0.017 (95% credible interval -0.05 to 0.10) for computerised speech and language therapy compared with usual care, resulting in an incremental cost-effectiveness ratio of £42,686 per quality-adjusted life-year gained. For computerised speech and language therapy compared with attention control, the incremental cost-effectiveness ratio was £40,164 per quality-adjusted life-year gained. Subgroup analyses resulted in incremental cost-effectiveness ratios of £22,371 and £28,819 per quality-adjusted life-year gained for computerised speech and language therapy compared with usual care in participants with mild or moderate word-finding difficulty, respectively. Using EuroQol-5 Dimensions, five-level version, quality-of-life scores reported by carers on behalf of patients instead of scores derived from the unofficial accessible version of the EuroQol-5 Dimensions, five-level version, resulted in an incremental cost-effectiveness ratio of £28,819 for computerised speech and language therapy compared with usual care.

Conclusions

The computerised speech and language therapy intervention provided additional hours of speech and language therapy to people with persistent aphasia at a low cost. Computerised speech and language therapy led to significant improvements in word-finding ability that were maintained irrespective of time post stroke. Word-finding improvements did not generalise to conversation or participant perceptions of communication participation and quality of life. Cost-effectiveness results suggest that computerised speech and language therapy is unlikely to be considered cost-effective for the whole population investigated, given typical current National Institute for Health and Care Excellence cost-effectiveness thresholds. Subgroup analyses are prone to increased uncertainty; however, the intervention may be more cost-effective for people with mild and moderate word-finding difficulties.

The computerised speech and language therapy intervention can be offered as part of speech and language therapy provision to achieve repetitive word-finding practice and improve the ability to find new words. This research suggests that it is important to use words of personal relevance in therapy owing to limited generalisation to untreated words, and that there needs to be a focus on supporting

the generalisation of new words into functional use. Further research recommendations in order of priority include:

1. investigating ways to assist with generalisation of newly learned vocabulary into use in conversation and other functional communication contexts
2. identification of what was practised, and whether or not all of the exercises set up by the speech and language therapist were used
3. exploration of further cost and time efficiencies
4. exploration of who to target the intervention towards
5. implementation of an optimised computerised speech and language therapy approach as part of NHS speech and language therapy provision
6. validation of the accessible variant of the EuroQol-5 Dimensions, five-level version.

Trial registration

This trial is registered as ISRCTN68798818.

Funding

This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 24, No. 19. See the NIHR Journals Library website for further project information. The Tavistock Trust for Aphasia provided additional support to enable people in the control groups to experience the intervention after the trial had ended.

Chapter 1 Introduction

Post-stroke aphasia

Aphasia is the most common language disorder acquired post stroke, affecting a person's ability to speak, read, write or understand language. One-third of stroke survivors are affected by aphasia and 30–43% of them will remain significantly affected in the long term.¹ Aphasia has an impact on everyday activity including the ability to have conversations, make telephone calls, listen to the radio, write letters, e-mails and text messages and read for pleasure, information or work. It therefore restricts the ability to carry out pre-stroke roles at work, in the family and in the community, often leading to withdrawal from participation in usual activities and reduced social networks. These changes affect the wellbeing of both the person with aphasia and their family/carer, with increased frustration, misunderstandings and breakdown of/strain on relationships. Consequently, people with aphasia are highly susceptible to depression.²

Evidence for speech and language therapy

Speech and language therapy is provided for people with aphasia. It aims to improve the language impairment and the ability to communicate and participate in daily activities. Neuroplasticity is the process by which the brain can form new connections and pathways to enable a person to relearn a skill, such as language, previously controlled by an area of the brain now affected by the stroke. Impairment-focused speech and language therapy aims to promote experience-dependent neuroplasticity for language. Key principles of experience-dependent neuroplasticity underpinning therapy according to Kleim and Jones³ include 'use it or lose it', 'use it and improve it', specificity matters (the nature of the therapy dictates the nature of the plasticity, i.e. you get better at doing what you practice doing), salience matters (therapy must be meaningful to induce plasticity) and repetition matters (induction of plasticity requires sufficient repetition).

In a Cochrane review of aphasia therapy for people with aphasia post stroke, Brady *et al.*⁴ showed that speech and language therapy was more effective than no treatment, resulting in clinically and statistically significant benefits to patients' functional communication, reading, writing and expressive language. They found no evidence of one type of therapy being superior to another and there is no current evidence of long-term effects of therapy.⁴ Medical instability, fatigue and confusion may reduce full engagement with language therapy in the early weeks post stroke, reducing the opportunity for some people to engage in therapy; however, some stroke survivors may tolerate speech and language therapy later post stroke.

Traditionally, it was thought that language recovery can reach a 'plateau' after 6 months or more, leading to speech and language therapy services being delivered predominantly in the first few months after stroke. However, in a systematic review of 21 randomised controlled trials (RCTs), Allen *et al.*⁵ found evidence to support the provision of speech and language therapy for more than 6 months after the onset of post-stroke aphasia. In a recently published RCT of speech and language therapy for aphasia in the chronic stage post stroke (> 6 months), Breitenstein *et al.*⁶ randomised 156 patients with aphasia to either speech and language therapy or waiting list control in 18 rehabilitation centres in Germany and showed significant statistical and clinical functional language (activity) benefit of speech and language therapy. The therapy was delivered in clinical settings for ≥ 10 hours per week for at least 3 weeks (minimum dose of 30 hours) and combined one-to-one speech and language therapy, group therapy with a speech and language therapist (SLT) and self-managed computer therapy or pencil-and-paper linguistic exercises prescribed by a SLT.⁶

Delivery of therapy for persistent aphasia: the clinical problem

Despite evidence of benefit, treatment of aphasia that persists beyond the first few months after stroke is often not available through NHS services,⁷ as ongoing therapy is costly through face-to-face speech and language therapy and places greater demands on already-limited resources, which are predominantly targeted in the earlier months following stroke. If people with aphasia are to be able to reach their recovery potential, lower-cost options for the support of repetitive practice in the longer term are urgently needed to enable access to therapy at a time that is best for each individual, and for as long as the individual is able to benefit.

Potential solutions to increasing the amount of tailored therapy delivered for persistent aphasia without increasing demand on speech and language therapy resources

There is evidence that non-speech and language therapy professionals can be employed successfully to support therapy activity and the Cochrane review found little indication of a difference in the effectiveness of therapy facilitated by volunteers trained by a SLT and the effectiveness of therapy delivered by a SLT directly.^{4,8,9} Computer technology can also provide the potential for supporting treatment in the long term. Computer programs developed for the treatment of aphasia have been reported to be useful in the provision of targeted language practice and provide opportunities for independent home practice as part of a self-managed approach to maximise repetitive practice,¹⁰⁻¹² improving outcomes for reading, spelling, word-finding and expressive language.^{11,13-15} Computer-based tasks can be tailored to the individual's needs, accounting for personal context and language ability levels, potentially helping to motivate independent practice. Self-managed practice schedules can also account for personal needs, such as fatigue, ability to travel and fitting practice around other commitments. Bespoke software and applications (apps) are available for self-managed aphasia therapy practice. In addition to personal computers (PCs), Stark and Warburton¹⁶ showed the feasibility of using iPads (Apple Inc., Cupertino, CA, USA) to deliver app-based aphasia therapy. In a systematic review of computer therapy for aphasia, Zheng *et al.*¹⁷ concluded that therapy delivered using a computer is more effective than no therapy, and potentially as effective as therapy delivered by a SLT. The meta-analysis carried out in the Cochrane review⁴ similarly concluded that there was no evidence of a difference between therapy delivered on a computer and one-to-one therapy from a SLT. Both studies acknowledge the low quality of this evidence, with only five small RCTs (the largest having 55 participants) reported to date (2016). Only one of these five studies considered impact on functional communication, indicating that the majority of the evidence is for impairment-based outcomes rather than functional or activity-based outcomes. No a priori sample size calculations were reported. Computer-based services for long-term management of aphasia therapy have the potential to provide a low-cost therapy option. However, the actual cost-effectiveness has not been investigated until recently in the Clinical and cost-effectiveness of aphasia computer treatment versus usual stimulation or attention control long term post-stroke (CACTUS) pilot study.¹⁸ There is therefore a pressing need for fully powered, well-designed RCTs of both the clinical effectiveness and the cost-effectiveness of self-managed computer aphasia therapy approaches for aphasia.

Intervention aimed at addressing delivery of long-term speech and language therapy for persistent aphasia investigated in the Big CACTUS trial: self-managed, computer aphasia therapy approach for persistent aphasia

The computerised approach to long-term aphasia therapy used in the Big CACTUS project [see *Chapter 2, Self-managed computerised therapy intervention for word-finding (computerised speech and language therapy)*] combines current theory and evidence underpinning language therapy with practical considerations for treatment delivery. As word-finding is a common difficulty for people with aphasia,

this intervention focuses on the treatment of word-finding specifically. The approach has four main components:

1. Access to specialist speech and language therapy software.
2. Skills of a qualified SLT used to select individually targeted therapy exercises (specificity) to practice retrieval of words of personal relevance to each individual with aphasia (salience).
3. Regular self-managed practice of the therapy exercises (20–30 minutes per day over 6 months is recommended as shown to be manageable in the pilot study).¹⁸
4. Volunteers/speech and language therapy assistants (SLTAs) support use of the computer exercises and generalisation of the newly acquired vocabulary in conversation.¹⁹

The intervention was predominantly focused on improving the word-finding impairment, with support from volunteers or SLTAs designed to assist with generalisation of words learned to functional use.

The computer software package used within this trial is called StepByStep© version 5.0 (Steps Consulting Ltd, Acton Turville, UK),²⁰ and is marketed by Steps Consulting Ltd at a cost of £250 for an individual lifetime licence to be purchased for a patient to put on their own computer, or £550 for a clinician licence owned by the NHS and installed on an NHS computer. The Stroke Association funded the initial development of the first version of StepByStep in the early 2000s. Since that time, it has been iteratively developed and marketed by Steps Consulting Ltd. Version 5 was used in the Big CACTUS trial. The approach described above is based on a similar approach used in therapy by Steps Consulting Ltd as an independent therapy provider. The software is intended to be tailored by SLTs with practice supported by a non-SLT specialist, often a carer or relative. The approach has been adapted for use in the NHS, particularly recognising that not everyone who accesses NHS services has a carer/relative able to provide support and therefore a training programme for volunteers and SLTAs was developed. Steps Consulting Ltd was not a collaborator on the project and therefore was not involved in project design, delivery or analysis. It did, however, support therapists on the trial with software use as it would for any therapist who purchased the software independently of the trial. For rationale for the choice of StepByStep, see *Chapter 2, Self-managed computerised therapy intervention for word-finding (computerised speech and language therapy)*.

We carried out a pilot study evaluating the approach described above with 34 people with persistent aphasia.²¹ They were randomly assigned to using the computer therapy approach or usual long-term care (most frequently this was social support). On average, people with aphasia practised their speech exercises on the computer independently for 25 hours over 5 months. The therapy significantly improved people's ability to use spoken words when compared with usual care (UC). The mean improvement in word-finding was 19.8% [95% confidence interval (CI) 4.4% to 35.2%; $p = 0.014$]. The results indicated that self-managed computer therapy supported by volunteers (total of 4 hours' support over 5 months on average) could help people with chronic aphasia to continue to practise, improving their vocabulary and confidence when speaking.²¹ Patients and carers found it an acceptable alternative to face-to-face therapy.²² The pilot study also showed, through qualitative interviews, that self-managed computer therapy could potentially improve the quality of life of people with persistent aphasia,²² at a relatively low cost to the NHS and society, but that a full economic evaluation with a larger sample was still required to reduce uncertainty in estimates of cost-effectiveness.¹⁸

Research rationale and objectives

The literature shows that people with persistent aphasia can improve their communication with sufficient amounts of speech and language therapy. This can be difficult to provide face to face with limited resources. Consequently, the use of specialist computer software for self-managed repetitive practice with volunteer support has been explored as a potential option for the provision of effective

speech and language therapy to people with persistent aphasia, providing the opportunity for people with aphasia to receive greater quantities of therapy over a longer period than would be possible face to face. The aim of the Big CACTUS trial was to provide definitive evidence of whether or not targeted, speech and language impairment based therapy intervention for word-finding through self-managed computer exercises for persisting post-stroke aphasia in addition to currently available face-to-face speech and language therapy was clinically effective and cost-effective, when compared with currently available speech and language therapy alone. This built on the successful 3-year Research for Patient Benefit-funded pilot RCT conducted by this team, which informed possible effects, measures, feasibility, recruitment rates, compliance, cost-effectiveness analysis and a power calculation. Results demonstrating feasibility were published by Palmer *et al.*²¹

The World Health Organization recommends use of the *International Classification of Functioning, Disability and Health* (ICF)²³ to describe and evaluate the impact of health problems on a person's life. As the intervention in Big CACTUS predominantly targeted word-finding impairment anticipating carry-over to functional activity, both impairment and activity were relevant to evaluate, along with participation. The first three research objectives therefore sought to identify the effect of the Big CACTUS approach to self-managed computer treatment for persisting aphasia supplementing UC (for a definition of UC in this trial, see *Chapter 2, Usual-care control group*), compared with UC alone or activity/attention control (AC) plus UC on the ICF dimensions of impairment, activity and participation:

The main research objectives were to:

1. establish whether or not self-managed computerised speech and language therapy (CSLT) for word-finding increases the ability of people with aphasia to retrieve vocabulary of personal importance (impairment)
2. establish whether or not self-managed CSLT for word-finding improves functional communication ability in conversation (activity)
3. investigate whether or not patients receiving self-managed CSLT perceive greater changes in social participation in daily activities and quality of life (participation)
4. establish whether or not self-managed CSLT is cost-effective for persistent aphasia post stroke
5. identify whether or not any effects of the intervention are evident 12 months after therapy has begun.

Additional research objectives include investigating the generalisation of treatment to retrieval of untreated words (impairment); the generalisation of treated words to functional use in conversation; carers' perception of communication effectiveness (participation) and the impact on carers' quality of life; and identification of any possible negative effects. Consistent with our objectives, we selected assessments to measure impairment, activity and participation (see *Chapter 2* for details). Since our trial was designed, the international aphasia research community has developed a consensus statement about the importance of measuring these ICF dimensions and has identified recommended measures.²⁴

Patient, carer and public involvement

The CACTUS pilot study had a strong patient, carer and public involvement (PCPI) group, which acted as an independent advisory group made up of people with aphasia and carers. This group was refreshed at the beginning of the Big CACTUS trial, with four members (two people with aphasia and two carers) providing continuity from the pilot study and three new members (two people with aphasia and one carer) joining the group to provide a fresh perspective. Members were recruited via a stroke patient and public involvement (PPI) database held by the University of Sheffield. The work of the group has been reported throughout this report using the Guidance for Reporting Involvement of Patients and the Public 2 (GRIPP2).

The aims of the group were to:

- facilitate the recruitment and inclusion of people with aphasia in the trial
- ensure that trial materials and processes were accessible to people with aphasia
- ensure that the interventions and trial procedures were appropriate and manageable for people with aphasia
- ensure that dissemination of trial results reached a broad audience in accessible formats.

The group met with members of the research team and was predominantly facilitated by the chief investigator, Rebecca Palmer, who is a qualified SLT. Involving people with language difficulties presents additional challenges for PCPI collaboration. Standard supportive techniques used included having an aphasia-friendly agenda at each meeting (see study documentation at www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#/), using practical activities with pictures and keywords to support discussions, and, for key decisions, members of the research team worked with each member with aphasia independently to facilitate inclusion of their perspectives. The clinically trained members of the research team worked with those members with the most limited language ability. The group members met when their advice or help with decision-making or production of materials was required and therefore the frequency of meetings varied. Fifteen meetings took place during the project.

The level of involvement was collaborative. Activities included development of the plain English summary for the original application; assisting with style and content of information sheets for people with different severities of aphasia; assisting with the design and evaluation of the consent support tool used in recruitment; advising on recruitment; informing the design of an aphasia-friendly adapted EuroQol-5 Dimensions, five-level version (EQ-5D-5L); making decisions about the materials used in the AC group; identifying key facts from the results that need to be particularly highlighted for people with aphasia; driving the dissemination plans; and stimulating and co-producing a film and accessible booklet of the trial results. The contribution of the PCPI group has been further detailed in *Chapter 2*. The impact of the PCPI activity has been considered in the discussion section (see *Chapter 6*). The group was awarded a PCPI prize at the UK Stroke Forum 2016 in recognition of their contribution to the Big CACTUS project.

In addition, a person with aphasia and his wife were members of the Trial Steering Committee (TSC).

Chapter 2 Methods

This report is concordant with the Consolidated Standards of Reporting Trials (CONSORT) guidelines for individually randomised parallel-group trials²⁵ (including non-pharmacological treatments,²⁶ pragmatic²⁷ and harms²⁸ extensions) and the Template for Intervention Description and Replication (TIDieR) checklist.²⁹

Trial design

The trial used a pragmatic, superiority, individually randomised, single-blind (blinded outcome assessors), parallel-group, randomised controlled adjunct trial design. All participants received UC, and outcomes were compared for people with persistent aphasia 4 months or more post stroke who were randomly allocated to one of the following groups:

- UC
- self-managed CSLT in addition to UC
- AC in addition to UC.

Participants were randomised to one of the three groups using a 1 : 1 : 1 allocation ratio. Randomisation was stratified by the NHS speech and language therapy department (site) providing the interventions and severity of the patient's word-finding difficulty.

Important changes to methods after trial commencement

The planned sample size was 285 participants. We extended the recruitment period by 1 month to increase recruitment as it was slightly short of the target owing to the recruitment rate slowing down towards the end of the recruitment period. The reduction in recruitment rate is likely to be caused by the fact that recruitment was from current and past patient lists and by the end of recruitment past patient lists had been exhausted. After an extra month of recruitment, the trial team, in discussion with the Trial Steering Committee and the trial statistician, made a decision to stop recruitment at 278 participants, seven participants short of the planned target, as the withdrawal rate was lower (9%) than that estimated in the sample size calculation (15%) and therefore sufficient numbers had been recruited to address research objectives with the intended statistical power.

The original funded trial did not include measurement of fidelity, beyond that of participant adherence to the interventions. Fidelity measurement of additional components of CSLT (i.e. delivery by SLTs and support from volunteers or SLTAs) was later funded by the Stroke Association as part of a postgraduate research [Doctor of Philosophy (PhD)] fellowship awarded to one of the Big CACTUS research associates, Madeleine Harrison, who was supervised by two of the Big CACTUS collaborators, Rebecca Palmer and Cindy Cooper. The methods for the fidelity assessment were added to the Big CACTUS protocol version 4.0, dated 17 July 2015, 8 months after recruitment commenced.

Important protocol changes since version 1.0

In the original protocol, the co-primary outcome 'conversation' was measured using two assessments. These were the Therapy Outcome Measures (TOMs) and the number of target words used in conversation. The trial was originally powered on only the TOMs and the other co-primary outcome, 'word-finding ability'. It was not possible to power the trial also on 'number of words used in conversation' as there was no prior information to inform a sample size calculation. Prior to the trial starting and publication of the protocol,

it was recognised that there was not adequate information to combine the two measures of the co-primary outcome of 'conversation' into one outcome. Therefore the validated, published measure, TOMs, was kept as the co-primary outcome and number of words used in conversation was moved to be a secondary outcome before recruitment began.

An additional 'per-protocol' definition for intervention use was added for clarity during the recruitment period but before data lock and analysis: 'across at least a 4 month period will be considered per protocol'.

An inter-rater and intrarater reliability testing protocol for raters of the practice videos in the primary outcome measure was added prior to evaluation of the videoed conversations.

Details of key subgroup analyses were added during development of the statistical analysis plan (SAP) prior to data lock: severity of word-finding difficulty, length of time post stroke and baseline comprehension ability.

Participants and eligibility criteria

Participants were included if:

1. They were aged ≥ 18 years.
2. They were diagnosed with stroke(s). Studies often limit inclusion to first stroke. As this is a pragmatic trial, and patients with multiple strokes are typically treated, inclusion was not limited to patients with a first stroke.
3. Their onset of stroke was at least 4 months prior to randomisation (to ensure that aphasia was persistent).
4. They had been diagnosed with aphasia, subsequent to stroke, confirmed by a trained SLT.
5. They scored 5–43 out of 48 on the Comprehensive Aphasia Test (CAT) Naming Objects³⁰ (mild is 31–43, moderate is 18–30 and severe is 5–17; participants scoring < 5 were excluded as the pilot study showed no benefit for participants who were able to retrieve $< 10\%$ of words).
6. They were able to perform a simple matching task in StepByStep with at least 50% accuracy (score of at least 5/10; this was a pragmatic method that may be used clinically to confirm sufficient visual and cognitive ability to use the computer exercises).
7. They were able to repeat at least 50% of words in a simple word-repetition task in the StepByStep program (score of at least 5/10). Significant difficulty with repeating words is an indication of apraxia of speech, which would require a different intervention.

Participants were excluded from the trial if they:

1. Had another premorbid speech and language disorder caused by a neurological deficit other than stroke. A formal diagnosis could be reported by the participant or relatives and confirmed by the recruiting SLT.
2. Required treatment for a language other than English (as the software is currently only available in English).
3. Were currently using the StepByStep computer program or other computer speech therapy aimed at word retrieval/naming to avoid similarity between groups.

Eligibility of providers

NHS speech and language therapy departments were eligible to participate if they routinely provided community services for people with aphasia post stroke. Treating clinicians were eligible if they were

qualified SLTs with experience of treating post-stroke aphasia. Speech and language therapy or generic rehabilitation assistants were eligible if they routinely carried out work under the supervision of a qualified SLT. Services were invited to use volunteers to provide the same support as assistants if they routinely used volunteers to support speech and language therapy work.

Settings and locations where the data were collected

Participant identification

Participants were recruited from 21 speech and language therapy departments in 20 NHS trusts (see *Appendix 1*) across the UK, from both current and past patient records, speech and language therapy colleagues and contacts with longer-term voluntary support groups. Identification from past patient records and voluntary support groups was aimed at including participants who may have finished their speech and language therapy intervention based on currently available services. However, these potential participants would be eligible for additional/extended therapy if the Big CACTUS approach to providing more therapy through self-managed computer exercise was to be implemented in the future if found to be clinically effective and cost-effective. Speech and language therapy departments agreeing to participate in the project were asked to identify potential participants. Potential participants unknown to speech and language therapy departments and voluntary groups, who found the trial on the website, were able to self-present by contacting the central trial team who put them in touch with their nearest local NHS speech and language therapy departments in the trial where possible. Potential participants (those identified as having had a stroke, and a diagnosis of aphasia, at least 4 months post stroke, aged ≥ 18 years) were contacted by the research SLT in each participating NHS trust. This person was a member of the local clinical team who was appointed to take responsibility for the running of the trial and the trial intervention in their NHS trust. The participant was sent project summary information and followed up by a telephone call 1 to 2 weeks later to establish whether or not they were interested in the trial. If they were, the research SLT made an appointment to visit them at home. All trial procedures, including recruitment, intervention and outcome measures, were conducted in the participants' own homes.

A screening log was completed by the therapist who identified potentially eligible patients from patient records, speech and language therapy colleagues or voluntary groups. Data recorded and sent back to the clinical trials research unit (CTRU) included unidentifiable information including initials, gender and age. The reason, if given, for not arranging an appointment was recorded.

Screening for eligibility

At the first visit to the potential participant, before providing detailed trial information, the research SLT determined whether or not the person was eligible. They requested verbal consent to undertake the naming test of the CAT,³⁰ which is used in routine practice and can establish the severity of a person's word-finding deficit. If the word-finding score was < 5 (10%) or > 43 (90%) (out of 48), an explanation was given to the patient that this type of computer therapy was not suitable for them. If they were still interested in computer-based therapies, they were directed to the aphasia software finder (www.aphasiasoftwarefinder.org; accessed 20 June 2018) developed to help patients with aphasia to identify software that would be most suitable for them. If the potential participant had eligible word-finding scores, the research SLT asked them to attempt a simple matching task on the computer to confirm ability to see the screen and perform simple tasks.

Recruitment

The level of support required to enable a person with aphasia to provide informed consent is dependent on the severity and profile of the aphasia. Considerable attention was given to the recruitment of participants with aphasia to ensure informed consent. In order to provide information in a format consistent with each individual's language ability, the Consent Support Tool was used.³¹ The research SLT at each site requested verbal consent from the potential participant to carry out part A of the Consent Support Tool (language screening test of 5–10 minutes). The result indicated which style of information would best

support their understanding of the trial. Three different styles of information sheet were available to enable as many participants as possible to make their own decision regarding whether or not to consent to participation in this trial. The consent support tool, carer information sheets and consent forms can be accessed online (see www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#/related-articles), as can aphasia-friendly/accessible information sheets and videos (see www.sheffield.ac.uk/scharr/sections/dts/ctru/bigcactus). Patient information sheet 1 was in large font with keywords emboldened for those who could understand written paragraphs. Patient information sheet 2 was for those who could read simple sentences but not full paragraphs. It followed standard aphasia-friendly principles with one idea presented per page in short simple sentences of large font. Keywords were emboldened and each idea was represented by a picture. Patient information sheet 3 was for those who could understand with significant support. Each idea/sentence was presented on a Microsoft PowerPoint® (Microsoft Corporation, Redmond, WA, USA) slide with simple text, keywords emboldened and picture support. Research SLTs were instructed to present each point in turn, read aloud to the potential participant and support with gestures, objects and drawings. The next sentence was then presented. All participants were given sufficient time to consider their participation before informed consent was taken by the research SLT. Participants providing their own informed consent were asked to sign an aphasia-friendly consent form.

The Consent Support Tool also identified individuals who did not have the mental capacity to consent for themselves owing to the severity of their language difficulty (those with severe aphasia who find it difficult to understand information, even with the support of adapted/pictorial information formats). These individuals receive speech and language therapy treatment in practice and may benefit from the trial intervention; therefore, it was important to include them. These potential participants were shown a short video clip of the computer program being used and of someone being assessed to show what was involved. If potential participants with severe aphasia who lacked capacity to consent indicated an interest, a relative (in Scotland, the person's legal representative or nearest relative) was asked to read the full information sheet and a covering letter detailing their responsibility, and to sign a carer declaration that they believed that their relative wished to take part (in Scotland, they were asked to sign a consent form). At the request of the PCPI group, all patients were given a copy of either the standard information sheet or the aphasia-friendly information booklet and a picture summary on one side of A4 paper.

For those participants with a carer, the carer was asked if they were willing to complete outcome measures related to their own quality of life and perception of their relative's communication ability. They were provided with the carer information sheet and were asked to sign a consent form.

Interventions

Following the TIDieR checklist,²⁹ the three trial interventions are described below.

Usual-care control group

Usual care for people with long-standing aphasia following stroke varies across the country in terms of type, frequency and length of provision, and is dependent on available resources in each locality. To accurately describe UC provided to people with aphasia > 4 months post stroke, patients, carers and therapists were asked what therapy they had received in the 3 months before they were randomised. Therapy notes were then consulted to record the dates of therapy sessions, therapy goals, length of sessions, personnel providing therapy and the mode of therapy delivery. The UC received prior to randomisation is reported here based on the 244 participants who were at least 4 months post stroke at the start of the observation period. These data can be considered as a trial result. However, as describing UC was not a research objective, and as the data describe one of the trial interventions, the information has been reported in this section.

What?

Of the people with aphasia at least 4 months post stroke in the 3 months prior to randomisation, 40% were in receipt of speech and language therapy and 60% were not. People with aphasia were more likely to receive therapy than not if they were < 12 months post stroke, but they were less likely to receive therapy if they were > 12 months post stroke. A lower proportion of people with mild word-finding difficulties (32%) received therapy than those with moderate (52%) and severe (40%) word-finding difficulties. Forty-eight per cent of the participants attended voluntary or social support groups. Twenty-three per cent of participants were in receipt of therapy but did not attend voluntary or social support groups, 32% attended support groups but did not receive therapy, 29% had neither and 16% had both.

Therapy aims or goals recorded for each therapy session in speech and language therapy notes were analysed using a quantitative content analysis by two SLTs (RP and HW). The goal categories, descriptions and examples in *Table 1* show the range of speech and language therapy activity forming UC. Approximately 50% of therapy goals in UC focused on rehabilitation of the language impairment. Twenty-three per cent of goals focused on enabling the person to communicate, often with compensatory strategies. Time was also spent on providing support, in addition to assessment and reviewing progress. More than 5 years after stroke, enabling goals were more prevalent than rehabilitation goals.

TABLE 1 Goals of therapy in UC

Goal category (level 1) and goal description (level 2)	Example (as described in the patient notes from which data were collected)	Number of goals	Percentage of goals
Assessment	<i>Assess higher level language functions</i>	44	4.8
Review	<i>Review progress made in therapy</i>	49	5.2
Rehabilitation (improving impairment)		4628	49.8
Comprehension	<i>Improve auditory comprehension</i>	21	2.3
Expressive language	<i>To produce longer/more complete verbal sentences</i>	87	9.4
Intelligibility	<i>Clearer speech</i>	15	1.6
Money skills	<i>Money handling skills</i>	14	1.5
Number skills	<i>Number recognition</i>	10	1.1
Phonological skills	<i>Phonological therapy</i>	32	3.5
Reading	<i>Identify functional written words</i>	81	8.8
Semantic skills	<i>Semantic categorisation of concrete items</i>	44	4.7
Word-finding	<i>To be able to find words in conversation with more ease</i>	107	11.5
Writing	<i>To be able to write short clear emails</i>	51	5.5
Enabling		211	22.7
Augmentative and alternative communication	<i>Functional communication using low tech AAC</i>	29	3.1
Conversation support	<i>Supported conversation using technology</i>	82	8.8
Participation in social conversation/ activities	<i>Speak more fluently with golf friends</i>	18	1.9
Total communication strategies	<i>Alternative ways to get message across</i>	19	2.1
Using everyday technology	<i>Use of spell check</i>	40	4.3
Word-finding/self-cueing strategies	<i>Functional and compensatory strategies for word-finding</i>	23	2.5

continued

TABLE 1 Goals of therapy in UC (continued)

Goal category (level 1) and goal description (level 2)	Example (as described in the patient notes from which data were collected)	Number of goals	Percentage of goals
Supportive		36	3.9
Emotional support	<i>Exploring loss and gain</i>	8	0.9
Improve mood	<i>To improve mood</i>	1	0.1
Increase confidence in communicating	<i>To improve confidence in talking in group setting</i>	6	0.7
Managing frustration	<i>Frustration levels</i>	1	0.1
Providing information	<i>To advise patient and family about impact and recovery from aphasia</i>	13	1.4
Support communication with other professionals/form completion	<i>Form filling support</i>	3	0.3
Support for family	<i>Communication support for family</i>	1	0.1
Vocational support	<i>Attend 'fit for work' interview</i>	3	0.3
Activity to support therapy		39	4.2
Discussing discharge	<i>Discharge planning</i>	5	0.5
Expert patient training	<i>Expert patient training</i>	2	0.2
Goal-setting	<i>To set goals for occupational therapy and speech therapy</i>	16	1.7
Handover	<i>Handover to new SLT</i>	4	0.4
Liaison with other staff/family	<i>Liaison with social worker</i>	4	0.4
	<i>To demonstrate laptop comprehension tasks to family</i>		
Preparing/monitoring homework	<i>Set up home exercises</i>	2	0.2
Therapy planning	<i>Establish motivation for therapy</i>	6	0.7
Insufficient information		86	9.3
Goal not sufficiently described	<i>Activity practice</i>	74	8.0
	<i>To achieve 90% on tasks</i>		
No goal recorded		12	1.3
Not communication therapy		1	0.1
Total		928	100.0

AAC, augmentative and alternative communication.

Who?

Usual speech and language therapy was predominantly provided by qualified SLTs at Agenda for Change bands 6 and 7. Some therapy sessions were provided by SLTAs on Agenda for Change bands 2 and 3.

How?

Usual speech and language therapy was predominantly provided face to face, with 87% of sessions delivered one to one and 12% of sessions delivered to a group of patients. Overall, five telephone calls were recorded as being used to provide therapeutic intervention and one instance of the use of telehealth was recorded.

Where?

Therapy was predominantly provided in patients' own homes when one to one, and in outpatient/ community health-care settings when provided in a group.

When and how?

The median therapy time received was 5 hours and 20 minutes, delivered in 1-hour sessions once every 2 weeks (median averages). The total time and number of sessions reduced with length of time post stroke, from 8 hours delivered in 1-hour sessions 0.67 times per week for people between 4 and 6 months post stroke to only 2 hours and 45 minutes delivered in 45-minute sessions once per month for people > 5 years post stroke (median averages).

Tailoring

Usual-care goals were tailored depending on a participant's interests and clinical decisions regarding their needs.

Modifications

As this was a pragmatic trial, UC varied between sites and participants according to usual practices. No attempt was made to standardise the UC received. Consequently, there were no planned modifications to UC during the trial.

Fidelity/measurement of how well usual care was delivered

Amounts of therapy time were recorded for UC throughout the trial to assess whether or not this stayed constant between trial groups. Fidelity to UC is reported in the trial results [see *Chapter 4, Usual-care speech and language therapy offered (fidelity/adherence to provision of usual care)*].

Self-managed computerised therapy intervention for word-finding (computerised speech and language therapy)

Why?

The intervention targeted word retrieval as it is one of the challenges most frequently experienced by people with aphasia, restricting their communication. The intervention was designed by SLTs specialising in aphasia and use of computer software for its treatment. The components of the intervention were designed to incorporate key factors that neuroplasticity principles and research suggest positively influence aphasia therapy outcomes combined with practical considerations: exercises tailored to the difficulty experienced by the individual with aphasia (specificity); content of therapy tailored to personal interests (salience); use of computer software to enable independent practice and therefore increased amounts of practice for a duration longer than that achievable through face-to-face therapy alone; and practical support and motivation for use of the software.

The four key components of the intervention are summarised below:

1. StepByStep software (version 5) – specialist aphasia software designed by SLTs and commercially available.
2. Qualified SLT assessment of the participant's language profile to tailor computer exercises using StepByStep so that they target the specific language deficit identified. Creation of exercises using target words of personal relevance to the participant.
3. Daily independent word-finding practice with the tailored computer exercises by the participant for 6 months.
4. Volunteer/SLTA support to enhance adherence to the computer exercises and to encourage transfer of new words into functional daily situations.

The TIDieR items 'what?', 'who?', 'how?', 'where?', 'when?', 'how much?' and 'tailoring' are described within each of the four components of the intervention in the following sections.

StepByStep software

What?

StepByStep software was chosen as it focuses on word-finding, allows for exercises to be tailored to individual need, enables personalisation through the addition of photographs (e.g. their spouse's) and provides feedback to the person with aphasia on practice frequency and duration and progress to aid motivation for repetitive practice. All of these features support the principles of experience-dependent neuroplasticity. The software was purchased by each NHS trust and provided to participants randomised to the computer therapy group of the trial. If participants had their own laptop/desktop or tablet computer, a home licence was installed by the SLT (for additional information on devices and microphone recommendations, refer to the Therapy manual: Big CACTUS StepByStep computer therapy approach at www.sheffield.ac.uk/scharr/sections/dts/ctru/bigcactus). If the participant did not own their own device, or their device could not run the software, a laptop or tablet computer owned by the NHS trust was loaned with a clinician licence of the software installed (for more than one user) for a period of 6 months. Participants with a home licence installed on their own computer kept this after the 6-month period. The combination of loaning devices for set periods of time and provision of home-user licences enabled equity of intervention provision, and reflected pragmatic decisions required for delivery of such interventions in clinical practice.

Qualified speech and language therapist assessment, tailoring of exercises and monitoring

Who?

A qualified SLT (one at each site) with experience of treating aphasia post stroke as part of the clinical team at that site.

What and tailoring

The SLT-tailored computer exercises to the individual using 100 words of personal relevance chosen by the participant. The word sets were standardised to 100 for each participant to allow sufficient words to maintain interest and motivation to practice for the long intervention period (up to 6 months). A meta-analysis of numbers of words used in word-finding therapy and outcome showed that people with all severities of aphasia could manage large word sets.³² There is a large bank of photographs within the StepByStep computer program from which to select personally relevant vocabulary and if something extra was required (e.g. a picture of a grandchild or favourite football team) it was photographed digitally and added by the SLT. The computer software²⁰ enabled the SLT to select exercises using these words; the exercises follow steps in the therapy process that the therapist would take if delivering word-finding therapy face to face. The SLTs based the selection of exercises on language skills demonstrated in the initial language assessments and therefore ensured that word-finding cues were useful and exercises were set at a level of difficulty with which the participant could experience success before moving on to more challenging exercises (see Therapy manual: Big CACTUS StepByStep computer therapy approach for the NHS, for instructions on how to modify exercises according to assessment results, at www.sheffield.ac.uk/scharr/sections/dts/ctru/bigcactus). The SLT provided an initial demonstration of the software exercises and spent time checking that the individual was able to use the software and monitoring the appropriateness of the tailored exercises. The SLT also reviewed the need for additional pieces of hardware, such as tracker balls, in order to make it physically possible for participants to use the computer.

Where?

The SLT carried out both assessment of and introduction to the computer exercises in the participants' own homes.

Regular self-managed practice

How, where, when and how much?

The participant was then asked to work through the exercises on the computer, with the aim that they would practise each day for 20–30 minutes. Participants were given a 6-month period to work through the therapy material on the computer at home and to practise using the new vocabulary in their daily lives. Practice with the computer for a minimum of 20 minutes three times per week at home on average across at least a 4-month period was considered per protocol. This accounted for periods of illness and holiday expected to occur in a 6-month period. The amount of practice was captured automatically by the computer program. Those participants who had the software installed on their own computers were not prevented from continuing to practise if they wished (with no prescribed support) following the 6-month supported intervention time. If computers were loaned, they were taken back after 6 months or when a new participant needed to borrow a computer (as permanent loan of equipment would be unusual in practice).

Volunteer or speech and language therapy assistant support with treatment adherence and carry-over into daily activity

Who?

To support use of the computer exercises, the SLT provided training to local volunteers who already had a working relationship with the speech and language therapy department (based in NHS trusts, local voluntary organisations or student SLTs) or SLTAs based in their department. This variation aimed to allow for consistency with the current mechanisms for providing therapy support in each NHS trust.

What, when, where, and how much?

The training programme and instruction book developed and evaluated during the pilot study was used (see Big CACTUS volunteer/assistant handbook at www.sheffield.ac.uk/scharr/sections/dts/ctru/bigcactus). The volunteer was asked to visit the participant at home for a minimum of 4 hours (once a month for 1 hour, or every 2 weeks for half an hour to suit the participant), carrying out the following tasks:

- provide technical assistance
- observe and encourage use of computer exercises
- check results and discuss difficulties
- assist the participant to move on to harder tasks in the therapy process pre-programmed by the SLT
- encourage the use of new words in everyday situations through conversation and discussions with family about how to encourage use
- set up new vocabulary sets if all 100 words had been completed.

Further advice provided to the volunteers/assistants on how to support the participant is detailed in the volunteer handbook (see www.sheffield.ac.uk/scharr/sections/dts/ctru/bigcactus). The participants were able to contact the volunteer/SLTA by telephone for technical advice on computer use between planned visits if necessary. The volunteer/assistant was introduced to the participant by the SLT on a joint home visit, and the volunteer was shown the exercises that had been set up. After each planned visit to the participant, the volunteer/assistant completed a feedback form for the SLT on what they did in the session and any issues/questions. The volunteer could contact the SLT by e-mail or telephone between support sessions to report any concerns/difficulties.

The computer intervention was delivered in addition to UC (see *Usual-care control group*).

Modifications

In response to feedback from the first four therapists providing the intervention to their first participants, the handbook was modified to explain that not all available cues needed to be tailored, only those assessed

as being useful for the individual. Provision was also made for therapists to provide the words to the participant over more than one session, to add cues for a subset of words and to review the usefulness of the cue before adding to all 100 words.

Fidelity/measurement of how well the intervention was delivered

The aim was to measure the effectiveness of the intervention as it would be delivered in clinical practice. The SLTs delivering the intervention attended 1 day of training on how to use the StepByStep software; training was provided by SLTs in the central Big CACTUS team based on training available to SLTs from Steps Consulting Ltd. They received a therapy manual (see www.sheffield.ac.uk/scharr/sections/dts/ctru/bigcactus).

The volunteers/assistants and SLTs were to support participants' adherence to computer practice as part of the intervention. However, no additional strategies were used to maintain or improve fidelity that could not be used in routine clinical practice.

Fidelity measures of the four key components of the intervention are outlined in the following sections.

StepByStep software

The proportion of participants allocated to the computer therapy group who had access to the StepByStep therapy software (the coverage of the intervention) was indicated by the completion of forms by the SLT to confirm that access had been given.

Assessment and tailoring of software by the speech and language therapist

Following training on how to set up the software and deliver the intervention, quality of delivery of the intervention was evaluated using a quiz of SLT knowledge about the intervention and a therapy planning form that SLTs completed each time they tailored the StepByStep software that captured their reasoning for the prompts and cues selected. SLTs delivering the intervention completed the quiz 5, 10 and 15 months after randomising their first participant. In addition, SLTs delivering the intervention were asked to complete therapy planning forms for an independent SLT to judge the extent to which exercises chosen were consistent with the language assessment results. For each SLT delivering the intervention, one therapy planning form was appraised by a StepByStep approach expert (RP, the Big CACTUS chief investigator, author of the therapy manual and a SLT experienced in delivering the StepByStep approach in clinical practice). Scores of 2 (reasonable rationale for tailored steps), 1 (partial rationale) or 0 (no or inexplicable rationale) were used. The time spent on each of the activities involved in delivering the intervention described in the manual was recorded by the SLTs delivering the intervention.

Practice of exercises by the participant

Adherence to exercise practice on the computer was captured automatically by the software and the total practice time was reported over the 6-month intervention period and compared with predefined per-protocol definitions (see *Chapter 3, Per-protocol sets*).

Volunteer/assistant support

The volunteers/SLTAs kept logs of the amount of time spent with each participant, including the number of sessions, duration of each session and session content.

The original Big CACTUS protocol funded by the Health Technology Assessment programme did not incorporate fidelity assessment. Measurement of fidelity to all components of the intervention was managed by a Big CACTUS research associate (MH) under the supervision of the chief investigator (RP) and co-investigator (CC) as part of a PhD funded by the Stroke Association. The results of the fidelity assessment described above are reported in the clinical results section of this report (see *Chapter 4, Fidelity to computerised speech and language therapy: adherence to practice and quality of intervention delivery*). A more-detailed evaluation of the fidelity to the intervention described within this trial will be available on completion of the PhD.

Attention control group

Why?

To control for the potential impact of elements of the computer intervention, which, of themselves, do not provide or require specific speech and language intervention.

What?

Participants were provided with generalised activities to carry out and general attention in addition to UC. On allocation to this group, the SLT conducting baseline assessments provided books of standard puzzles that could be purchased from most supermarkets or from high-street shops. Each book contained enough activities for one to be carried out each day for at least 1 month. Examples of puzzles include 'spot the difference', noughts and crosses, and word searches. The PCPI group advised on types of puzzle book that may be of interest and practical factors, such as size of the text.

Who?

The SLT provided age-appropriate puzzle books that matched the participant's linguistic and cognitive ability as indicated by the baseline assessments. Puzzle books were colour coded into levels of easy, medium and hard by the clinicians on the research team centrally with support from the PCPI group and a leaflet was provided to give SLTs guidance on skills required for each level.

Who, how, where, when and how much?

A member of the research team contacted the participant or their carer by telephone or e-mail (whichever was preferred by the participant) once a month for the duration of the 6-month intervention period to mimic the attention provided by volunteers in the intervention group. Participants were asked if they were enjoying the activities, how many they managed to do at home, whether or not they would like a new puzzle book sent to them for the coming month and whether or not it needed to be the same level of difficulty, or easier or harder. The participants also had access to contact details of the research team to enable them to ask for easier or harder books at any time if necessary, mimicking the access to the volunteers/SLTAs and type of attention available in CSLT.

Modifications

No modifications were made to AC during the course of the trial.

Fidelity/measurement of how well the intervention was delivered

The number of puzzle books sent out and the number of contacts made by the research team were used as a proxy measure of adherence to AC. A puzzle book was sent out if a participant or carer reported completing the previous one. A minimum of six puzzle books and four contacts was used as a measure of adherence to the intervention.

Outcomes

Primary outcome measures

Research objective 1: to establish whether or not self-managed computerised speech and language therapy for word-finding increases the ability of people with aphasia to use vocabulary of personal importance (impairment)

The change in word-finding ability, between baseline and 6 months, of words personally relevant to the participant was measured by a picture-naming task (100 words with a maximum of 2 points each).

The word-finding score was expressed as a percentage of the total score, and change in the percentage 6 months from baseline was calculated. This is a measure of the change in the impairment and was considered to be of interest to SLTs as it indicates whether or not word-finding treatment for persistent aphasia (i.e. beyond the acute and subacute phase) is effective for improving retrieval of words.

The pictures were presented within an assessment module of the StepByStep programme by the research SLT at baseline at the NHS trust from which the participant was recruited.

The research SLT was trained on how to score the word-finding test and was provided with written instructions in the outcome measure therapists' handbook (see www.sheffield.ac.uk/scharr/sections/dts/ctru/bigcactus). The research SLT was given two videos of people with aphasia carrying out word-finding tests for practice and this was reviewed by a speech and language therapy trainer from the central project team who provided feedback to the research therapist. The same test was performed 6 months after randomisation by an outcome measure therapist (a qualified SLT). This therapist was trained to score the test through webinar training from the central trial team. The scores derived from the two practice videos were compared with the scores of the research SLT in the same NHS trust to check for inter-rater reliability between raters within each trust. When there was a discrepancy, the research SLT and outcome measure therapist were encouraged to discuss differences and rate a third video.

Research objective 2: to establish whether or not self-managed computerised speech and language therapy for word-finding improves functional communication ability in conversation (activity)

Change in functional communication between baseline and 6 months was measured by blinded ratings of 10-minute video-recorded conversations between a SLT (research SLT at baseline and outcome measures SLT at outcome) and participants, using the activity scale of the TOMs.³³ Conversations were structured around topics of personal relevance to the participants by the SLT performing baseline measures, based on the 100 words they selected. The same conversation topic guide was followed by SLTs performing outcome measures. Independent SLTs blinded to treatment allocation and measurement time point rated the videoed conversations at the project co-ordinating centre. This measure of functional communication ability was used to indicate whether or not the word-finding intervention had any impact on the ability to communicate in conversation. TOMs were chosen as a primary outcome measure of this ability as they have been standardised, shown to be reliable and have been used to rate videoed conversations in a previous RCT of aphasia intervention [Assessing Communication Therapy in the North West (ACT NoW)⁸] with good reliability.

A benchmarking session using the TOMs was conducted with potential raters to get consensus on the application of the TOMs in this project, followed by inter-rater and intrarater reliability tests at least 6 weeks apart using 10 practice videos. Scoring instructions were provided following consensus during the benchmarking session (see *Appendix 2*). The consensus was that pairs of videos were easier to rate for each individual rather than isolated videos for each participant. The 14 raters selected for final rating of all participant videos had intrarater reliability of at least 70% (7/10) of practice videos rated within 0.5 between rating at time 1 and time 2, and inter-rater reliability of at least 70% (14/20) of videos rated within 0.5 of the median scores from all raters at both time points. In total, 86% (240) of the 20 ratings made by each of the 14 reviewers (total 280 ratings) were within 0.5 of the median score, and 88% (123/140 ratings) were within 0.5 between time 1 and time 2. A slight upwards trend was noted in scoring between time 1 and time 2; therefore, the pairs were presented in random order. For further detail of the process for selection of TOMs raters and the scoring procedure, see *Appendix 2*.

Key secondary outcome measure

Research objective 3: to investigate whether or not patients receiving self-managed computerised speech and language therapy perceive greater changes in social participation in daily activities and quality of life (participation)

Improvement in patient perception of communication between baseline and 6 months was measured using the Communication Outcome after Stroke (COAST) scale, a patient-reported measure of communication, participation and quality of life validated for evaluating speech and language therapy interventions in the Health Technology Assessment ACT NoW project.⁸ This measure was used to provide SLTs with quantitative information on participant perceptions of the effects of the intervention

on their life to complement the qualitative information collected through patient interviews in the pilot study. The COAST was administered face to face by the research SLT at each participating NHS trust at baseline and by outcome measures therapists at 6 months.

Secondary outcome measures

Research objective 4: to establish whether or not self-managed computerised speech and language therapy is cost-effective for persistent aphasia post stroke

A cost-utility analysis was undertaken from a NHS and Personal Social Services (PSS) perspective. The cost-effectiveness outcome was the incremental cost-effectiveness ratio (ICER), where effectiveness was measured in quality-adjusted life-years (QALYs). The incremental analysis included all three of the trial groups. Resource costs were estimated for patients, including intervention software and hardware, and SLT and assistant input time, combined with standard costing sources. Volunteer time was also recorded and costed for inclusion in a supplementary analysis taking a broader perspective. SLTs were asked to complete therapy activity forms for each contact with each participant, detailing their Agenda for Change pay band, time spent setting up the computer therapy or in face-to-face support, or support/training of the volunteer/assistant, and travel mileage. Assistants and volunteers were also asked to complete activity forms with information on their Agenda for Change pay grade if applicable, time spent face to face or indirectly with the participant, activity conducted with the participant and mileage.

The EQ-5D-5L questionnaire was administered at all time points and combined with standard valuation sources to measure QALYs gained in each treatment group. An accessible version of the EQ-5D-5L designed by the PCPI group for people with aphasia was completed by participants. An accessible version of the EuroQol-5 Dimensions, three-level version (EQ-5D-3L), had been tested in the pilot study.¹⁸ The carers (if available) completed the standard version by proxy. Carers also completed the EQ-5D-5L for themselves. For more detail about the use of EQ-5D-5L in this trial, see *Chapter 5, Health-related quality of life*.

Information on cost-effectiveness was important to inform commissioners of speech and language therapy services as well as providers to assist with decisions regarding funding such an intervention.

Research objective 5: to identify whether or not any effects of the intervention are evident 12 months after therapy has begun

Evidence of treatment effect was measured by repeating all outcome measures at 9 and 12 months from baseline in addition to the primary end point of 6 months. The 9-month time point was included as an interim measure as withdrawal from the trial was found to increase over time in the pilot study. Follow-up measurements were important to provide information to SLTs, commissioners and providers about the long-term impact of the intervention.

Additional secondary outcome measures

The first primary outcome measure identified improvement in the ability to retrieve words practised in therapy and the second primary outcome measure identified any improvement in conversation using the standard descriptors provided by the activity scale of the TOMs³³ to try to detect generalisation of any impairment level improvement to the level of activity. As an intermediate and potentially more sensitive measure of generalisation, use of words practised in therapy in the context of conversation was measured by two research members of the central trial team who watched each videoed conversation and identified how many words practised in therapy were used during the structured conversation on related topics. The researchers ticked the word on a checklist if it was heard and scored the total number of practice words heard. Intrarater and inter-rater reliability was established for the researchers. They both rated the same set of 10 videos twice, a minimum of 6 weeks apart. Inter-rater reliability was

METHODS

80% (8/10) at both time 1 and time 2. Intrarater reliability was 100% (10/10) for rater Kathryn McKellar and 90% (9/10) for rater Ellen Bradley. The researchers were blind to the time point at which each video was made.

Generalisation of treatment to retrieval of untreated words was measured using the object naming test from the CAT.³⁰ This measure was used to show whether or not generalisation occurs from treated to untreated words. This is important to SLTs so they know whether or not careful selection of vocabulary for their treatment of word-finding is important.

Carer perception of communication effectiveness was measured using the Carer COAST (CaCOAST).³⁴

The last five items of the CaCOAST and responses to the Care-related Quality of Life instrument (CarerQoL) were collected to indicate any impact of the intervention on the carers' quality of life.

As self-managed computer use for speech and language therapy is relatively new and, by nature, unsupervised, any negative effects specifically felt to be related to computer use were collected using a negative effects questionnaire (see www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#/related-articles) that was sent to the participants in the CSLT group every month. This asked participants whether the computer practice made them feel overtired, affected their eyes, gave them headaches or made them feel anxious or worried. They were also asked to list any other problems experienced as a consequence of using the computer therapy. In addition, adverse events (AEs) and serious adverse events (SAEs) were reported by the research therapists on discovery or following a telephone call to the participant or carer at 6, 9 and 12 months (see *Appendix 3* for definitions of AEs and SAEs for the population in this trial).

For the method for calculating scores for all of the above outcome measures, see *Chapter 3, Computation of summary outcome scores for analysis*.

Staff training for delivery of outcome measures

All research SLTs attended a 1-day training course delivered by the central research team to ensure understanding of the protocol and trial procedures including how and when to administer outcome measures. It was understood that some participants may need more than one visit to complete the assessments. The flow of participants through the trial was the responsibility of the research SLTs (principal investigators) at each participating NHS trust. A flow diagram of activity was provided to assist with this (see *Appendix 4*).

Therapists responsible for carrying out assessments at all follow-up time points (6, 9 and 12 months) were trained in their responsibilities, the importance of blinding and how to administer outcome measures. Training was in small groups via a webinar delivered by the trial manager and the central team speech and language therapy researcher. It was recommended that they conduct each outcome measure visit within 1 month of each time point to ensure that outcome measure visits were spaced sufficiently to avoid presenting a burden to participants.

Collection of demographic data

Initial assessment was undertaken by the research SLT at each participating NHS trust. The initial assessment visit included collection of demographic data: aphasia type, age, gender, time post onset of stroke, and type and location of stroke (if known).

Recording of usual care

The research SLT at each NHS trust was asked to record UC provided to each participant throughout the trial to provide a description of UC and also to ensure that UC did not change once participants were randomised to the trial. UC included care provided by both the NHS and the voluntary sector. In addition, the research SLTs were asked to record UC provided in the 3 months prior to

randomisation from participant and carer reports and from SLT notes. For each session, they were asked to report the therapy goal, length of session, mode of delivery of session, Agenda for Change band of staff delivering the session and distance travelled. This information was sought every 3 months via telephone calls with the participants and carers and from inspection of participant notes if they were made aware that therapy had been received.

Sample size

The trial aimed to recruit a maximum of 285 participants (95 per group) across 20–24 speech and language therapy sites to preserve 90% power for a 5% two-sided test to address both co-primary objectives relating to word-finding of personally selected words and functional communication outcomes. We adjusted for a 15% drop-out rate observed in an external pilot trial.²¹ For the change in word-finding, based on consensus of the therapists on the trial team and the aphasia PPI group, we assumed a 10% mean difference as clinically worthwhile to detect and a standard deviation (SD) of 17.38% estimated from an external pilot trial²¹ based on an analysis-of-covariance model.

We inflated the sample sizes by 1.14 to account for the fact that the variance was estimated from a pilot trial.^{35,36} For the change in functional communication (TOMs activity scale), we sought an effect size of 0.45 of the SD as clinically worthwhile and a 0.5 correlation between baseline and outcome observed in the ACT NoW study (Professor Andy Vail, University of Manchester, 2013, personal communication).

For the change in COAST, a key secondary outcome measure, we sought a 7.2% clinically worthwhile effect to detect and a SD of 18% based on externally supplied data and we assumed a 0.5 correlation between baseline and outcome. For a sample size of 285 (95 per group), the trial had 83% power for the COAST. The observed overall drop-out rate was about 9%, versus the planned 15%; as a result, further recruitment was terminated at 278 participants because the trial had the desired statistical power to address co-primary and key secondary outcomes.

Interim analyses and stopping guidelines

The initial phase of the trial was conducted as an internal pilot trial and included clear criteria to inform decisions about progression and the feasibility of the full trial only. No interim analysis for efficacy, futility or stopping early for safety was planned. Data from the internal pilot are included in the final analysis.

The internal pilot trial was limited to six sites (> 25% of the total). However, during this phase we recruited and commenced set-up processes for all of the intended sites to avoid a delay in the event that the trial continued. In accordance with the guidance on progression rules for Health Technology Assessment internal pilot trials, the lag phase expected before recruitment reached the target rate was excluded. For the substantive study, the lag phase included the period for obtaining approvals, site recruitment and staff training. The progression criteria were reviewed 8 months from site set-up of the sixth site.

Based on recruitment rates from the previously published pilot study,²¹ we aimed to recruit participants at an average rate of one participant per site per month. At the end of the internal pilot trial phase, the six pilot trial sites had been recruiting for a minimum of 8 months. The progression of the trial was based on achieving the following criteria.

Numbers recruited

The overall target for the six sites was 36 participants. The overall progression target for numbers recruited from the six sites was 30. This was equivalent to the number recruited in total in our previous pilot study and enabled comparison with previous recruitment rates to confirm whether or not our projections for the substantive study were accurate. There was also information available from other sites that had completed set-up and started to recruit; therefore, we expected at least 40 participants to have been recruited by the end of the internal pilot phase in total.

Recruitment as a percentage of the full-study recruitment targets

At the end of the internal pilot trial, progression depended on having recruited 30 participants (i.e. 10% of the total population recruited from 25% of the sites; this was midway through the recruitment phase for these sites). If we achieved this number, we would be on track to recruit only 80% of the sample size within the study period. We would then have to bring in the additional four contingency sites included in the costs to raise the recruitment to the sample size. If we did not meet this number, it would indicate that the larger study was unlikely to be feasible.

Retention to first outcome measure time point at 6 months (primary outcome)

The sample size calculation was based on an attrition rate of 15% at 90% power for the co-primary outcomes. The progression criterion for retention was set to ensure a minimum power of 80%. This would be achievable with a retention rate of 65%, which would still ensure that the results were generalisable.

Identification and retention of volunteers

Sites could provide support to patients in the intervention group of the trial from paid SLTAs or volunteers. Use of volunteers was reviewed at the end of the internal pilot phase. Progression criteria for continued use of volunteer support were set at 80% of participants having been offered a volunteer and 70% of participants continuing to be supported by the same volunteer for their 6-month treatment period. If these progression criteria were not achieved, continuation of the study would be with paid assistant support only.

Summary

In summary, 8 months after set-up of the sixth site, our progression criteria indicating feasibility of the full trial were:

- recruitment of no fewer than 30 participants (10% of the target for the full trial)
- a minimum retention rate of 65%.

Randomisation and concealment

We used a centralised web-based randomisation system hosted by the Sheffield CTRU to randomise participants to one of the three treatment groups using a fixed 1 : 1 : 1 allocation ratio. The randomisation schedule was generated using stratified block randomisation with randomly ordered blocks of sizes three and six, stratified by centre and the severity of word-finding at baseline based on scores of the naming test of the CAT (mild, 31–43; moderate, 18–30; and severe, 5–17). Only the randomisation statistician knew about the block sizes and they were disclosed after the trial had finished. A Sheffield CTRU statistician independent of the trial conduct logged on to the randomisation system to specify the randomisation details and generated the randomisation schedule, which was retained within the system. The system offers restricted access such that research team members are granted access to particular functionalities depending on their roles in the trial.

The SLTs randomised participants in their homes, following baseline assessments, using the Sheffield CTRU web-based online randomisation system and disclosed the allocation to the participant. If no internet connection was available, the SLT would telephone the research team at the Sheffield CTRU, who randomised online and gave the allocation immediately over the telephone to the SLT to disclose to the participant.

Blinding

This was a single-blind trial recognising that participants and lead research SLTs at each participating NHS trust could not be blinded to participants' treatment allocation as participants would be provided with software personalised by the research SLT, or provided with puzzle books, or neither, depending on their allocation. However, all baseline and outcome measures were conducted blind to treatment allocation. The research SLTs conducting baseline assessments did this prior to randomisation. After randomisation, the research SLTs were unblinded to treatment allocation to be able to provide the correct intervention. Therefore, separate outcome measure SLTs were trained at each participating NHS trust to ensure that 6-, 9- and 12-month follow-up assessments were conducted by assessors who were blind to treatment allocation. The outcome assessors were SLTs with no previous involvement in the conduct of the trial. They were trained, via a webinar session run by the central team, on the importance of remaining unaware of the treatment allocation of the participants they would be assessing. Research SLTs were asked not to disclose baseline case report forms, not to openly discuss participants with colleagues in open-plan offices and to remind their participants not to discuss their activities on the trial with any other SLTs they may come into contact with as it was 'a secret'. When outcome assessors contacted participants and conducted their assessments, they were advised to remind participants that their activity on the trial was 'a secret'. It was possible that during a conversation with the participant or carer, outcome assessors could become unblinded by the participant or their carer. If this happened on the telephone, before the assessment took place, then the assessment was carried out by a different, blinded, outcome measure SLT. If unblinding happened at the end of the visit, when the assessment was complete, this was not classed as an unblinded assessment, as the actual assessment was carried out when the assessor was still blinded. However, in the event of unblinding of the SLT during an assessment, the next assessment was carried out by a different blinded assessor. All sites had a minimum of two trained speech and language therapy assessors who were blinded to the allocation to allow for unblinding issues.

If the treatment group allocation was disclosed during an assessment, then the outcome assessor would continue with the assessment but subsequently alert the research SLT (principal investigator) and complete an unblinding form. The unblinding form asked the assessor to record what they believed the participant's treatment allocation to be ('the suspected allocation'). In some instances, the assessor would guess the treatment allocation incorrectly, so the central team would report this as 'suspected unblinding' only. Descriptive summaries of circumstances surrounding the unblinding of SLTs' cases were recorded and instances of unblinding are reported in *Chapter 4, Reported cases of unblinding of outcome assessments*, to indicate the relative success of blinding procedures.

Video recordings of conversations at baseline and 6, 9 and 12 months were rated for the primary outcome measure by SLTs, using TOMs, who were independent of the trial team at the central trial site. Two researchers from the trial team rated the videos for the secondary outcome measure of number of treated words used in conversation. In addition to blinding to treatment allocation, blinding to time point was also maintained for the raters by presenting the videos to be rated in random order.

The trial statistician did not have access to any information that could reveal participants' treatment allocations during the trial and the randomisation statistician was not the same individual as the trial statistician who performed the analysis.

METHODS

Methods of analysis are described in the chapters that follow: statistical analysis in *Chapter 3* and health economic analysis in *Chapter 5*.

The full trial protocol version 5.0 is available online (see www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#).

Ethics approval was obtained from Leeds West NHS Research Ethics Committee (reference number 13/YH/0377) and Scotland A Research Ethics Committee (reference number 14/SS/0023).

Chapter 3 Statistical analysis methods

This chapter describes the statistical methods and principles used to analyse the trial to address the clinical effectiveness objectives [all main and additional research objectives except the objective referring to cost-effectiveness (number 4); see *Chapter 1, Research rationale and objectives*]. The SAP version 1.2 was written to conform to the International Conference on Harmonisation topic E9,³⁷ applicable standard operating procedures from the Sheffield CTRU and the trial protocol. The SAP was signed off before blinded or unblinded review of the data and is accessible online (see www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#). Post hoc analyses are explained with rationale. All analyses were undertaken using Stata[®] 15.1 (StataCorp LP, College Station, TX, USA).³⁸

Analysis populations

Modified intention-to-treat set

The modified intention-to-treat (mITT) population includes all participants for whom consent was obtained, for whom treatment was allocated as per the randomised list regardless of circumstances after randomisation and who had primary outcome data at the 6-month assessment. This was the primary analysis set; other sets described below are for sensitivity analysis.

Complete-case set

This includes participants with outcome data at a particular assessment. The set was used for subsidiary analysis of co-primary and key secondary outcomes at 9 and 12 months, and other secondary outcomes at different assessments, and for plotting mean profile response of participants over time across interventions.

Per-protocol sets

The goal here was to explore the effectiveness of the intervention among participants who adhered to key components of the intervention. Therefore, the per-protocol sets include participants for whom key components of the intervention were adhered to, including achieving the minimum amount of practice recommended (20–30 minutes per day) and having access to support up to and including at their 6-month assessment. Per-protocol classification relating to adherence to the intervention was undertaken for only the CSLT and AC groups.

Per-protocol definitions were developed using the following rationale.

The recommended practice time was 20–30 minutes per day over 6 months. It was calculated that 26 hours provided an indication of high levels of adherence to this recommendation for the CSLT group based on a minimum of 20 minutes three times per week for at least 4 months, accounting for periods of holiday, illness and other commitments. A lower total practice time of 10 hours was also considered to identify engagement with the computer intervention even if at a lower practice amount than recommended. Participants in the AC group were recommended to complete one puzzle each day. We estimated that, on average, there were enough puzzles in a book to last for 1 month so we used six puzzle books requested/sent as a proxy for adherence to carrying out regular puzzle book activities. The intervention descriptions recommend that support should be provided once per month in both the CSLT group and the AC group. Accepting that this is not always possible, four contacts was used as an indicator that support was being provided.

For all intervention groups, participants were excluded from the per-protocol analysis if outcome measures were assessed 14 days before or 31 days after the expected 6-month assessment or if they were randomised but failed to meet at least one inclusion criterion.

The four per-protocol sets in the CSLT and AC groups were matched as follows:

1. high levels of adherence to recommended practice – practised computer therapy for a minimum total of 26 hours (CSLT) or were sent at least six puzzle books (AC) within 6 months of randomisation (referred to as PP1 CSLT26 AC6)
2. adherence to practice – practised computer therapy for a minimum total of 10 hours (CSLT) or were sent at least six puzzle books (AC) within 6 months of randomisation (PP2 CSLT10 AC6)
3. high levels of adherence to recommended practice and provided with recommended support – practised computer therapy for a minimum total of 26 hours (CSLT) or were sent at least six puzzle books (AC) and contacted at least four times (if they wished) (in both the AC group and the CSLT group) within 6 months of randomisation (PP3 CSLT26 AC6_4)
4. adherence to practice and provided with recommended support – practised computer therapy for a minimum total of 10 hours (CSLT) or were sent at least six puzzle books (AC) and contacted at least four times (if they wished) (in both the AC group and the CSLT group) within 6 months of randomisation (PP4 CSLT10 AC6_4).

On 7 March 2018, the Trial Management Group discussed the final trial results in detail. It was noted that only a small proportion of participants in the AC group met the predefined per-protocol inclusion criteria. It was agreed that the predefined per-protocol proxy of being sent six puzzle books in 6 months was overambitious and inconsistent with expectations of per protocol for CSLT. To align the puzzle book per-protocol classification in the AC group with the CSLT group, the Trial Management Group requested post hoc per-protocol classification based on being sent at least four puzzle books. This will only change the per-protocol results for comparisons involving the AC group. We therefore modified the per-protocol classifications presented above for sensitivity analysis as follows:

1. high levels of adherence to recommended practice – practised computer therapy for a minimum total of 26 hours (CSLT) or were sent at least **four** puzzle books (AC) within 6 months of randomisation (PP1 CSLT26 AC)
2. adherence to practice – practised computer therapy for a minimum total of 10 hours (CSLT) and were sent at least **four** puzzle books (AC) within 6 months of randomisation (PP2 CSLT10 AC 4)
3. high levels of adherence to recommended practice and provided with recommended support – practised computer therapy for a minimum total of 26 hours (CSLT) or were sent at least **four** puzzle books (AC) and contacted at least four times (if they wished) (in both the AC group and the CSLT group) within 6 months of randomisation (PP3 CSLT26 AC4_4)
4. adherence to practice and provided with recommended support – practised computer therapy for a minimum total of 10 hours (CSLT) or were sent at least **four** puzzle books (AC) and contacted at least four times (if they wished) (in both the AC group and the CSLT group) within 6 months of randomisation (PP4 CSLT10 AC4_4).

Multiple imputation set

This includes all randomised participants excluding those who died prior to 6 months; it is for sensitivity analysis. Multiple imputation (MI) was conducted for only the co-primary outcomes (word-finding and functional communication) and key secondary outcome (COAST). The mean value of participants with available baseline data was used to impute missing baseline data during analysis for a few participants with follow-up outcome data but who had missing related baseline data for some reason. We adopted a strategy to inform the MI statistical model for imputing missing data:

1. Potential predictors of outcomes independent of the intervention were clinically prespecified in the SAP (see www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#).
2. The characteristics of completers (those meeting the mITT inclusion criteria) and all randomised participants (excluding those who died) were descriptively compared to explore predictors of missing data.
3. We graphically explored the association between baseline characteristics and outcomes of interest.

Based on these exploratory results, the intervention group appeared to be a mild predictor of missing data. Therefore, it was implausible to assume that the data were missing completely at random. Other measured predictors of missing data were unclear. As a result, we adopted the following strategy to impute missing data:

- Intervention group (UC, AC or CSLT), age, gender (male or female), the presence of a carer (yes or no), severity of word-finding (total score), severity of comprehension ability (total score) and baseline outcome measure under consideration were mandatory covariates in all MI statistical models.
- Co-primary and key secondary outcomes at baseline (covariates) and 6, 9 and 12 months were included in all MI statistical models.
- The longitudinal nature of the outcome data under consideration was accounted for in the MI statistical models using chained equations^{39,40} implemented via the Stata 15.1 *mi* command.

The multiple imputation using chained equations analysis was conducted and reported in accordance with the guidance provided by White *et al.*⁴¹ We chose the number of imputations ($n = 20$) based on the observed proportion of missing data.⁴¹

Linear interpolation set

For additional sensitivity analysis on the impact of missing data on the trial results, we used a linear interpolation model as a deterministic imputation approach. Where data were missing at an assessment (t_i) but valid data are available at previous (t_{i-1}) and future (t_{i+1}) assessments, the missing value was linearly interpolated by the formula:

$$Y_{\text{missing}} = Y_{i-1} + (Y_{i+1} - Y_{i-1}) \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}. \quad (1)$$

Safety set

Safety analysis relates to the evaluation of the intervention effect on AEs and SAEs, and negative effects of the computer therapy. This includes all randomised participants with informed consent and treatment allocation for analysis used the actual intervention received (treatment as received principle) based on available evidence, such as the number of books sent and computer therapy practice time. We also used treatment allocation as randomised for sensitivity analysis.

Statistical considerations

Approach to dealing with deaths after randomisation

We expected some deaths in this trial population during the trial. The research team discussed implications of deaths and approaches to handle them during analysis. The influence of the trial interventions on increasing risk of mortality was unanimously viewed to be very unlikely. In addition, the interpretation of imputed missing data, such as word-finding and functional communication for participants who died, was clinically challenging. As a result, the research team agreed not to impute data that were missing because of participant deaths. Therefore, deaths that happened prior to the 6-month assessment were excluded in any clinical effectiveness analysis but included in the safety analysis. This approach is consistent with related recommendations.⁴² We performed sensitivity analysis to explore the impact of data that were missing for other reasons, unrelated to death, on the bias of the results using the MI, linear interpolation and complete-case sets.

Hochberg multiple testing procedure

We used the Hochberg procedure to interpret the co-primary (word-finding and functional communication) and key secondary (COAST) outcome results at 6 months in order to control the chances of falsely declaring statistically significant results (at 5% nominal level).⁴³ There are two sources of multiple hypothesis testing at

6 months: multiple outcome measures (co-primary measures and a key secondary measure) and key multiple treatment comparisons (CSLT vs. UC and CSLT vs. AC). Figure 1 displays the interpretation strategy of the results in order to claim statistical significance and superiority of the intervention.

Computation of summary outcome scores for analysis

This section details the generation of outcome scores from measurement instruments and how related missing data were handled.

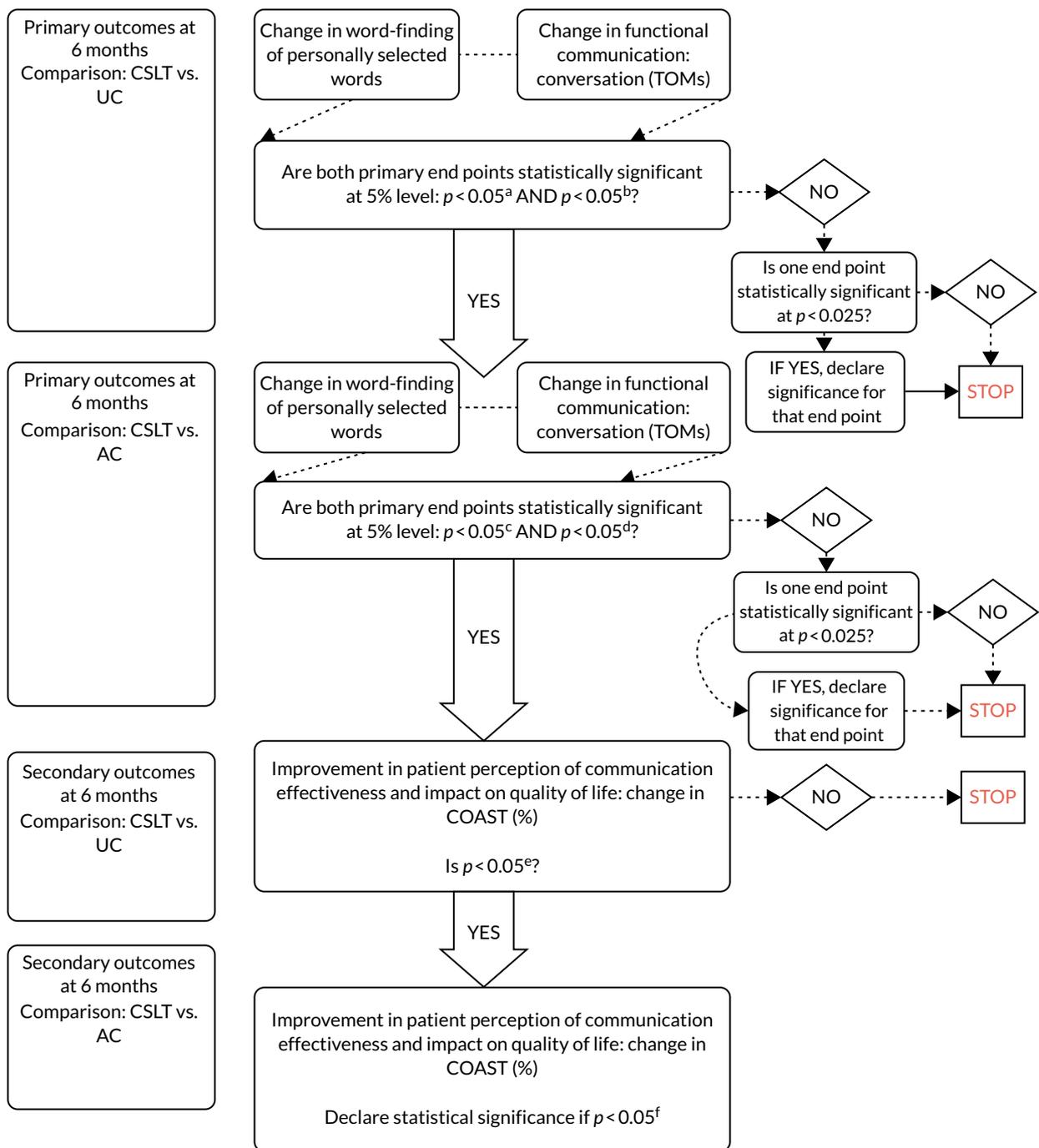


FIGURE 1 Interpretation of the Hochberg hierarchical sequential hypotheses testing strategy. Superscript letters correspond to footnotes in Table 8.

Classification of the severity of word-finding difficulty and comprehension ability

We categorised the severity of word-finding difficulty at baseline based on total scores from the CAT Naming Objects using a severity rating of mild (31–43), moderate (18–30) and severe (5–17). In addition, we generated a categorical comprehension ability variable at baseline based on the total scores from the CAT Comprehension of Spoken Sentences as follows:

- severe (0–8) – inconsistently understanding at the two information-carrying words level
- moderate (9–17) – consistently understanding at the two/three information-carrying words level/simple sentence structures but not complex sentence structures
- mild (18–26) – some understanding of complex sentence structures but not consistent
- within normal limits (27–32) based on CAT cut-off score for normal/aphasic.

Word-finding of personally selected words for treatment

As highlighted in *Chapter 2, Primary outcome measures*, we used the personal vocabulary naming test to assess word-finding ability based on 100 personally selected words for treatment. For each personally selected word, word-finding ability was assessed using the following scoring system: 0 for an incorrect or no response; 1 for a word named correctly after a delay of at least 5 seconds and/or for a self-correction; and 2 for a correct prompt answer within 5 seconds. This method was consistent with the scoring method used in the CAT Naming Objects. This scoring system yields a potential maximum score of 200. Although all participants were expected to be assessed based on 100 personally selected words, it was possible that some participants could have been assessed based on fewer than 100 words for some reason, such as fatigue. At least 70 words should have been assessed for an assessment to be considered valid. If fewer than 100 words but more than 70 words were assessed, the word-finding ability for the participant, k (Y_k), expressed as a percentage, was calculated based on the total score relative to the potential maximum score as:

$$Y_k = \frac{\sum_{i=1}^X \text{item}_i}{2X} \times 100, \quad (2)$$

where $i = \{1, 2, \dots, X\}$ is the picture item on the personal vocabulary naming test and X is the total number of personally selected words assessed.

Functional communication in conversation

The activity dimension of the TOMs instrument was used to assess functional communication in conversation. The rating is measured on a 6-point ordinal scale ranging from 0 (unable to communicate in any way) to 5 (communicates effectively in all situations), which allows scoring between ordinal descriptors such as 0.5, 1.5, 2.5, 3.5 and 4.5. Thus, the rating scale has 11 ordinal possibilities that can be treated as a continuous scale. There is a ceiling effect for participants who are able to communicate effectively in all situations at baseline with a TOMs rating of 5. We report the numbers and proportions of these participants across interventions.

Social participation and quality of life

The COAST patient-reported measure was used to assess self-perceived communication/social participation and impact on quality of life.³⁴ The measure has 20 items and each item is assessed on a rating scale of 0 to 4. Other responses ('not applicable', 'unclear' or 'no responses') are permitted. A procedure is then applied to compute a percentage score under a number of scenarios: all applicable and answered items, the existence of 'not applicable' items, the existence of 'unclear' or 'no response' items. We computed the overall percentage score using a validated algorithm as described by Bowen *et al.*⁴⁴

Carers' perception of patient communication effectiveness and of their own quality of life

The CaCOAST assesses carer perception of patients' communication effectiveness and impact on their quality of life.⁴⁵ The measure has 20 items; each item is assessed on a scale of 0 to 4, and a percentage summary measure is calculated. The CaCOAST was administered by the research therapists as one questionnaire. The first 15 items assess carer perception of patients' communication and the last five items assess the impact of the patients' communication difficulties on the carers' quality of life. Although the original validated scoring algorithm is based on all 20 items, we considered the first 15 items (CaCOAST 15) and the last five items (CaCOAST 5) separately to assess different questions. We therefore modified the scoring algorithm consistent with the validated scoring system using all 20 items⁴⁴ to compute the CaCOAST 15 (%) and CaCOAST 5 (%), based on the first 15 and last five items, respectively. These use the same scoring algorithm as described for the COAST but account for missing data as other aspects are uninformative ('not applicable' and 'unclear').

Generalisation of treatment to untreated words

The CAT Naming Objects consists of 24 pictures of words to assess word-finding. We used this to measure generalisation to untreated words. For each picture to be named, the following scoring system applies depending on a participant's response: 0 for an incorrect response, 1 for an accurate response after a delay of more than 5 seconds and 2 for an accurate and prompt answer. A total score ranging from 0 to 48 is then generated to assess word-finding of untreated words. Missing information (item level or all items) was possible owing to tiredness or being unable to complete the tests. For missing items, summary scores from the CAT Naming Objects were calculated assuming a conservative worst-case scenario: a 0 for a missing item score. No summary score was calculated if all items were missing.

Use of treated words in conversation

The use of treated vocabulary in the context of the conversation was assessed using a checklist of target words during ratings of videoed conversations at 6 months. Out of the 100 treated words, personally selected for treatment, the number of words retrieved during videoed conversations was counted (total score ranging from 0 to 100). A correct word retrieved was counted only once regardless of the number of times it was retrieved during the conversation.

EuroQol-5 Dimensions, five-level version utility and visual analogue scale

The approach described here applies to the carer, patient proxy and patient aphasia-friendly versions of the EQ-5D-5L. The EQ-5D-5L was used to assess health status and produces a single index value for health status for use in the calculation of QALYs to inform health economics evaluation of investigative interventions.⁴⁶ The instrument consists of a EQ-5D-5L descriptive system and a EQ-5D-5L visual analogue scale (VAS). The descriptive system has five dimensions assessing mobility, self-care, usual activity, pain/discomfort and anxiety. Each of these dimensions has five levels of severity, which participants were asked to select one of to best describe their health status 'today': no problems, slight problems, moderate problems, severe problems and extreme problems. Based on participants' responses from these five dimensions, a single index value was calculated. The index values are on a scale of 1 (full health) to 0 (state equivalent to dead) and health states considered to be worse than dead attain negative values (< 0). For the analysis of clinical effectiveness, the index values were estimated directly from the EQ-5D-5L data, as detailed by Devlin *et al.*⁴⁷ For the base-case economic analysis, index values were estimated using the approach currently recommended by the National Institute for Health and Care Excellence (NICE),⁴⁸ by converting EQ-5D-5L responses to utility scores using the UK cross-walk mapping algorithm developed by van Hout *et al.* in 2012.⁴⁹ The economic analysis tested the sensitivity of the cost-effectiveness results to the various techniques of estimating utility scores from EQ-5D-5L data, as described in *Chapter 5, Health-related quality of life*. As for the EQ-5D-5L VAS, participants were asked to rate how good or bad their health is 'today' on a scale of 0 (the worst health imaginable) to 100 (best health imaginable). The scores from this continuous scale assess change in overall self-rated health status.

Statistical analysis methods

Clinical effectiveness

For the co-primary and key secondary outcomes at 6 months, we used a multiple linear regression model adjusted for associated baseline outcome measures and fixed stratification factors: centre and the severity of word-finding (mild, moderate or severe). We expressed the maximum likelihood estimate of the intervention effect as the adjusted mean difference in change (MDC) between the CSLT and UC groups, and the CSLT and AC groups, with associated 95% CIs and *p*-values. The intervention effect between the AC and UC groups obtained via contrasts was for exploratory purposes only. We used a Hochberg procedure, as described in *Hochberg multiple testing procedure*, to control for the false-positive error rate for claiming the evidence of clinical effectiveness.

For sensitivity analysis, we used a multiple linear regression model adjusted for associated baseline outcome measures, fixed stratification factors (centre and the severity of word-finding), the length of time post stroke (continuous) and the location of stroke (yes or no) (middle cerebral artery, frontal lobe, parietal lobe and temporal lobe).

The long-term intervention effect at 9 and 12 months on the co-primary and key secondary outcomes, and other secondary continuous outcomes (at 6, 9 and 12 months), such as CaCOAST domains, word-finding of untreated words and word-finding of treated words used in conversation, was evaluated using a multiple linear regression model adjusted for associated baseline outcome measures and fixed stratification factors: centre and the severity of word-finding (mild, moderate and severe).

We used a multiple logistic regression model adjusted for fixed stratification factors (centre and the severity of word-finding) to explore the intervention effect on the proportion of participants achieving predefined clinical improvements of 5% and 10% in word-finding of both treated words (from personal vocabulary naming test) and untreated words (from CAT Naming Objects). The numbers and proportion of participants meeting each clinical improvement criterion are reported by intervention together with the odds ratio (OR) and associated 95% CIs and *p*-values. This was performed under two scenarios by considering (1) only participants with complete data and (2) all randomised participants, but assuming that those with missing data failed to achieve clinical improvement (worst-case scenario). Related results are presented in the statistical report version 1.0, which is available online as a supplementary appendix to Palmer *et al.*⁵⁰

Functional impact

For post hoc analysis at the request of the chief investigator following the disclosure of the results (as per predefined SAP), we explored the intervention effect on the proportion of participants who did and did not use treated words in conversation based on a 5% or 10% improvement in naming treated words (from personal vocabulary naming test) and functionally used at least 5 or 10 treated words (retrieved during videoed conversation). This analysis was requested to identify if any patients were able to retrieve newly learned words in functional contexts so that further investigation of characteristics of those who do and do not automatically use new words in context can be explored in further research to inform how others might be assisted to use new words functionally. We calculated the proportion of participants meeting each clinical improvement criterion by intervention group. The difference in proportions of participants achieving a 'clinical improvement' criterion between interventions was calculated, with associated 95% CIs estimated using the normal approximation to the binomial distribution without significance testing.

An additional post hoc analysis explored the proportion of participants who do or do not generalise word-finding to untreated words. A multiple logistic regression model adjusted for fixed stratification factors (centre and the severity of word-finding) and baseline measures were used to explore the intervention effect on the proportion of participants achieving clinical improvement of 5% and 10%

in the generalisation of word-finding to untreated words (from CAT Naming Objects) without significance testing.

Subgroup evaluation

We prespecified the following subgroups to explore potential heterogeneity in the intervention effect on the co-primary and key secondary outcomes at 6 months (SAP version 1.2; see www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#):

- the severity of word-finding difficulty – mild (31–43), moderate (18–30) and severe (5–17)
- comprehension ability – within normal limits (27–32), mild (18–26), moderate (9–17) and severe (0–8)
- the length of time post stroke – categorisation was quantile based because there was no existing literature to guide the clinical classification.

We performed subgroup analysis based on the mITT set. The number of participants and mean change in outcomes are reported stratified by the intervention received and subgroup category. We assessed effect modification between the intervention and subgroup using a multiple linear regression model that included an interaction term between the intervention and the subgroup of interest adjusted for baseline outcome measures and fixed stratification factors: centre and the severity of word-finding difficulty (mild, moderate and severe). We report the overall *p*-values from the interaction tests to explore the strength of evidence for heterogeneity of the intervention effect across subgroups. We use forest plots to present results and aid visual interpretation, showing the maximum likelihood estimate of the intervention effect (CSLT vs. UC and CSLT vs. AC), with associated 95% CIs stratified by subgroup category.

Safety evaluation

The primary analysis of safety outcomes was based on the safety set. We also performed sensitivity analysis on the safety outcomes using the treatment as randomised, as described in *Safety set*.

Negative effects of computer use in computerised speech and language therapy

The number and proportion of participants who experienced any perceived negative effects are summarised stratified by negative effect category: tiredness (fatigue), vision, headaches and anxiousness or worrisome. We calculated the total number of repeated events experienced by a participant per negative effect category and exposure (follow-up contributing to the 6-month data). We used a negative binomial regression model accounting for overdispersion and the exposure without significance testing to estimate the incidence rate (IR) with 95% CI.

Adverse events and serious adverse events

We calculated the number and proportion of participants who experienced any AEs or SAEs by intervention. For each participant, we calculated the exposure (trial follow-up) and the number of repeated AEs and SAEs. We used a negative binomial regression model to estimate the IR in each intervention and incidence rate ratio (IRR), with associated 95% CI accounting for overdispersion and the exposure without statistical significance testing.

Chapter 4 Clinical effectiveness results

Introduction

This chapter reports the results to address the clinical effectiveness research questions in conformity with the CONSORT guidelines for individually randomised parallel-group trials (2010)²⁵ (including non-pharmacological treatments,²⁶ pragmatic trials²⁷ and harms²⁸ extensions). The statistical analysis methods used and related trial outcomes are detailed in *Chapter 3* and the SAP is accessible online (www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#). *Chapter 6* discusses these results in detail together with other findings.

Statistical results

Screening and flow of trial participants

Between September 2014 and August 2016, 995 participants were screened and 818 were found to be appropriate for further assessment of eligibility across 20 NHS trusts in the UK (21 randomisation centres/speech and language therapy departments), predominantly via patient records (including word of mouth from speech and language therapy colleagues) and voluntary support groups. *Figure 2* shows the flow of participants from screening to randomisation including trial follow-up at 6, 9 and 12 months from randomisation. Of the 995 patients screened, 288 (28.9%) eligible participants were consented and 278 (27.9%) were randomised: UC, $n = 101$; AC, $n = 80$; and CSLT, $n = 97$.

Of the 278 patients randomised, eight (2.9%) died before their 6-month assessments: UC, $n = 4$; AC, $n = 1$; and CSLT, $n = 3$. For the remaining 270, 240 (88.9%) completed their 6-month assessments: 86, 71 and 83 in the UC, AC and CSLT groups, respectively. The proportions of participants who completed 6- and 9-month outcome assessments were very similar across interventions. However, the discontinuation rate at 12 months was slightly higher in the AC (23.8%) and the CSLT (23.7%) groups than in the UC (16.8%) group. Attrition reasons are stated in *Figure 2* and detailed in *Table 2*. The most common reasons were personal or family issues and being unhappy with the allocated trial intervention.

In total, 240 randomised participants were eligible for inclusion in the primary mITT analysis: 86 in the UC group, 71 in the AC group and 83 in the CSLT group. As for the secondary MI analysis, 270 randomised participants were eligible for inclusion: 97 in the UC group, 79 in the AC group and 94 in the CSLT group.

Characteristics of participating speech and language therapy departments in NHS trusts

The 21 speech and language therapy departments participating in the trial served a mixture of urban and rural populations. Ten sites were described by local principal investigators as predominantly rural, four as predominantly urban and seven as mixed rural and urban. Twenty-six eligible SLTs were trained to deliver the CSLT intervention in total. Sixteen sites chose to use only SLTAs to provide the support component of the intervention, three sites used only volunteers and two sites used both SLTAs and volunteers.

Baseline demographics and characteristics of randomised participants

Tables 3 and *4* show the demographics and characteristics of randomised participants stratified by intervention; 270 met the MI inclusion criteria and 240 met the mITT inclusion criteria ('completers'). Baseline characteristics of the eight participants who died before the 6-month outcome assessment are not reported as they were not included in the analysis because the association between the interventions and increased risk of mortality was viewed as extremely unlikely.

CLINICAL EFFECTIVENESS RESULTS

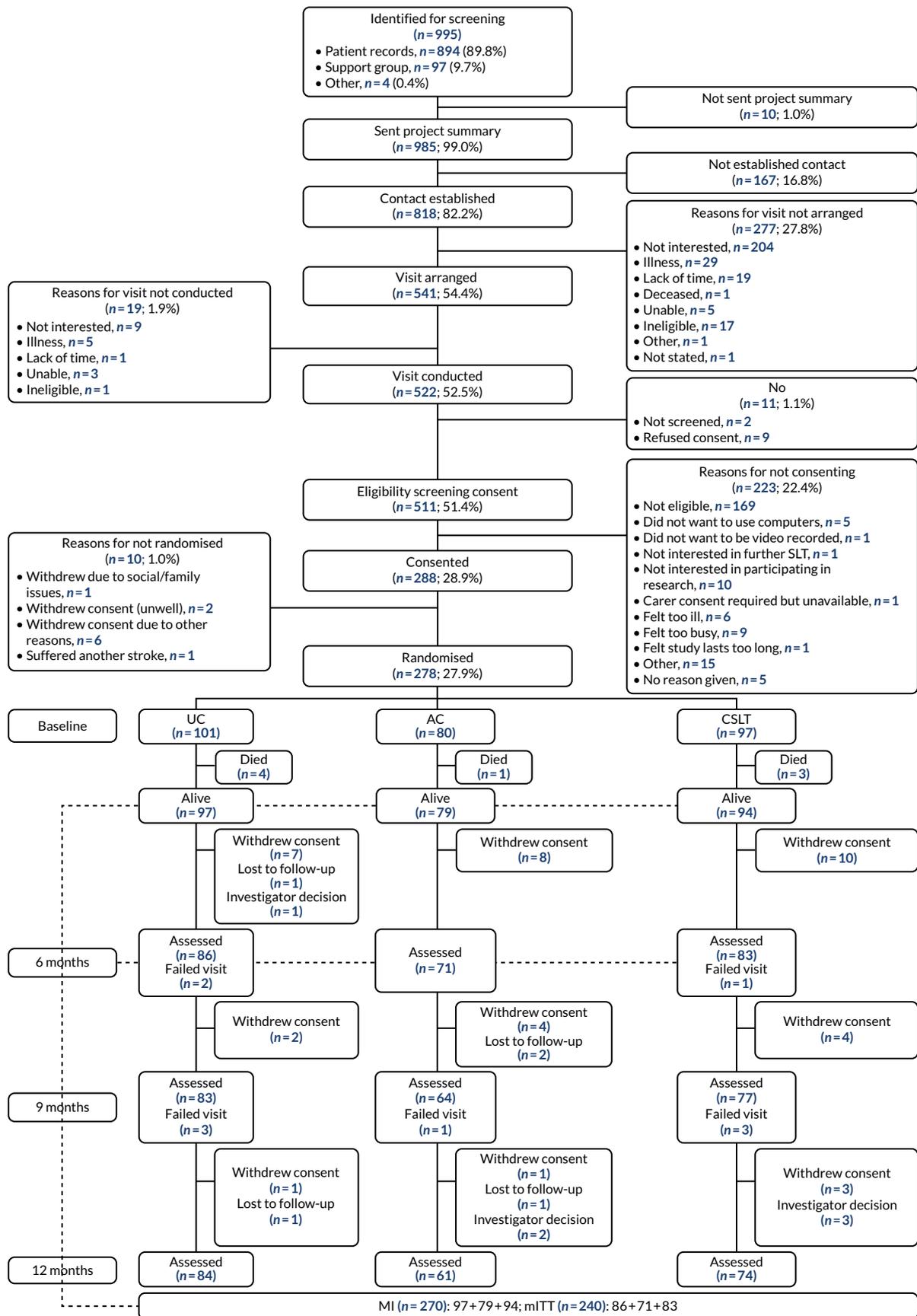


FIGURE 2 Trial participant flow chart. Adapted from Palmer *et al.*⁵⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

TABLE 2 Reasons for attrition

Discontinuation type	Reason for withdrawal	Number of participants			
		UC (n = 17)	AC (n = 19)	CSLT (n = 23)	Total (n = 59)
Death	N/A	4	1	3	8
Investigator decision	Personal/family issue	1	2	3	6
Lost to follow-up	N/A	2	3	0	5
Participant withdrew consent	Personal/family issue	5	4	6	15
	Unhappy with allocated trial group	4	3	4	11
	Unwilling to complete outcome measures	1	3	2	6
	Prefers not to say	0	1	1	2
	Moving out of the area	0	0	1	1
	Other (lost motivation and feeling unwell)	0	1	0	1
	Other (time commitment)	0	0	1	1
	Other	0	1 ^a	2 ^b	3

N/A, not applicable.

a In the AC group, one participant wanted to spend their time doing other activities that they enjoy more and feel are more beneficial.

b In the CSLT group, one participant struggled with voice recognition not working consistently, difficulties with the computer and frustration. One reported that they found it upsetting it does not do well ('It's not doing me any good') despite explaining that it is just the monitoring now to help with the study and she has done all the hard work'.

TABLE 3 Baseline demographic characteristics of randomised participants

Variable	MI population (N = 270)			mITT population (N = 240)		
	UC (n = 97)	AC (n = 79)	CSLT (n = 94)	UC (n = 86)	AC (n = 71)	CSLT (n = 83)
Site, n (%)						
Ayr	3 (3.1)	2 (2.5)	4 (4.3)	3 (3.5)	2 (2.8)	4 (4.8)
Belfast	5 (5.2)	3 (3.8)	3 (3.2)	5 (5.8)	3 (4.2)	3 (3.6)
Cambridgeshire	5 (5.2)	3 (3.8)	5 (5.3)	5 (5.8)	3 (4.2)	4 (4.8)
Cwm Taf	3 (3.1)	1 (1.3)	5 (5.3)	2 (2.3)	1 (1.4)	4 (4.8)
Derbyshire	6 (6.2)	5 (6.3)	5 (5.3)	6 (7.0)	5 (7.0)	5 (6.0)
Dorset	4 (4.1)	3 (3.8)	5 (5.3)	4 (4.7)	3 (4.2)	5 (6.0)
Glasgow	8 (8.2)	6 (7.6)	8 (8.5)	8 (9.3)	5 (7.0)	7 (8.4)
Hull	6 (6.2)	4 (5.1)	5 (5.3)	5 (5.8)	4 (5.6)	4 (4.8)
Newcastle	6 (6.2)	5 (6.3)	4 (4.3)	6 (7.0)	4 (5.6)	4 (4.8)
Norfolk	3 (3.1)	5 (6.3)	2 (2.1)	2 (2.3)	4 (5.6)	2 (2.4)
North Bedford ^a	3 (3.1)	3 (3.8)	4 (4.3)	3 (3.5)	3 (4.2)	4 (4.8)
North Lincolnshire	4 (4.1)	5 (6.3)	2 (2.1)	3 (3.5)	3 (4.2)	2 (2.4)
Northampton	5 (5.2)	4 (5.1)	6 (6.4)	5 (5.8)	4 (5.6)	6 (7.2)
Northern	5 (5.2)	4 (5.1)	3 (3.2)	5 (5.8)	4 (5.6)	3 (3.6)
Nottinghamshire	6 (6.2)	6 (7.6)	7 (7.4)	5 (5.8)	6 (8.5)	7 (8.4)

continued

TABLE 3 Baseline demographic characteristics of randomised participants (continued)

Variable	MI population (N = 270)			mITT population (N = 240)		
	UC (n = 97)	AC (n = 79)	CSLT (n = 94)	UC (n = 86)	AC (n = 71)	CSLT (n = 83)
Plymouth	3 (3.1)	1 (1.3)	3 (3.2)	3 (3.5)	1 (1.4)	2 (2.4)
Sheffield	5 (5.2)	4 (5.1)	6 (6.4)	5 (5.8)	4 (5.6)	6 (7.2)
Somerset	4 (4.1)	5 (6.3)	5 (5.3)	1 (1.2)	4 (5.6)	3 (3.6)
South Bedford ^b	6 (6.2)	5 (6.3)	4 (4.3)	5 (5.8)	4 (5.6)	2 (2.4)
Sunderland	3 (3.1)	2 (2.5)	5 (5.3)	2 (2.3)	2 (2.8)	4 (4.8)
Swansea	4 (4.1)	3 (3.8)	3 (3.2)	3 (3.5)	2 (2.8)	2 (2.4)
Gender, n (%)						
Male	60 (61.9)	49 (62.0)	55 (58.5)	54 (62.8)	44 (62.0)	47 (56.6)
Female	37 (38.1)	30 (38.0)	39 (41.5)	32 (37.2)	27 (38.0)	36 (43.4)
Age at consent (years)						
	n = 97	n = 79	n = 94	n = 86	n = 71	n = 83
Mean (SD)	65.6 (13.1)	64.8 (13.1)	65.6 (12.7)	64.9 (13.0)	63.8 (13.1)	64.9 (13.0)
Median (IQR)	66.6 (55.8–74.7)	66.2 (54.6–74.9)	66.1 (55.5–75.5)	66.5 (55.1–74.3)	65.1 (53.0–73.4)	64.7 (54.5–74.7)
Min., max.	23.1, 91.8	30.4, 88.7	34.1, 89.2	23.1, 89.6	30.4, 88.7	34.1, 89.2
CAT comprehension score^a						
	n = 97	n = 79	n = 94	n = 86	n = 71	n = 83
Mean (SD)	21.0 (6.0)	19.5 (7.2)	20.0 (7.0)	21.0 (5.9)	19.8 (7.0)	20.1 (7.3)
Median (IQR)	22.0 (17.0–26.0)	21.0 (14.0–25.0)	21.5 (15.0–26.0)	22.0 (17.0–26.0)	21.0 (14.0–26.0)	22.0 (14.0–26.0)
Min., max.	0.0, 30.0	1.0, 30.0	0.0, 32.0	0.0, 30.0	1.0, 30.0	0.0, 32.0
CAT comprehension severity,^a n (%)						
Severe	3 (3.1)	6 (7.6)	5 (5.3)	3 (3.5)	3 (4.2)	5 (6.0)
Moderate	24 (24.7)	24 (30.4)	29 (30.9)	20 (23.3)	24 (33.8)	26 (31.3)
Mild	50 (51.5)	36 (45.6)	43 (45.7)	46 (53.5)	31 (43.7)	35 (42.2)
Within normal limits	20 (20.6)	13 (16.5)	17 (18.1)	17 (19.8)	13 (18.3)	17 (20.5)
Severity of word-finding difficulty,^c n (%)						
Mild	40 (41.2)	38 (48.1)	41 (43.6)	35 (40.7)	35 (49.3)	36 (43.4)
Moderate	33 (34.0)	19 (24.1)	28 (29.8)	29 (33.7)	17 (23.9)	26 (31.3)
Severe	24 (24.7)	22 (27.8)	25 (26.6)	22 (25.6)	19 (26.8)	21 (25.3)
Type of aphasia, n (%)						
Anomic	39 (40.2)	22 (27.8)	35 (37.2)	33 (38.4)	19 (26.8)	33 (39.8)
Non-fluent (e.g. Broca's)	40 (41.2)	29 (36.7)	38 (40.4)	36 (41.9)	27 (38.0)	34 (41.0)
Mixed non-fluent	13 (13.4)	21 (26.6)	15 (16.0)	13 (15.1)	20 (28.2)	11 (13.3)
Fluent (e.g. Wernicke's)	5 (5.2)	7 (8.9)	6 (6.4)	4 (4.7)	5 (7.0)	5 (6.0)
Evidence of apraxia of speech, n (%)						
No	64 (66.0)	48 (60.8)	62 (66.0)	55 (64.0)	42 (59.2)	52 (62.7)
Yes	33 (34.0)	31 (39.2)	32 (34.0)	31 (36.0)	29 (40.8)	31 (37.3)
Type of stroke, n (%)						
Infarct	79 (81.4)	64 (81.0)	69 (73.4)	69 (80.2)	58 (81.7)	60 (72.3)
Haemorrhage	14 (14.4)	7 (8.9)	14 (14.9)	12 (14.0)	6 (8.5)	13 (15.7)
Not known	9 (9.3)	8 (10.1)	11 (11.7)	9 (10.5)	7 (9.9)	10 (12.0)

TABLE 3 Baseline demographic characteristics of randomised participants (continued)

Variable	MI population (N = 270)			mITT population (N = 240)		
	UC (n = 97)	AC (n = 79)	CSLT (n = 94)	UC (n = 86)	AC (n = 71)	CSLT (n = 83)
Location of stroke, n (%)						
Middle cerebral artery	47 (48.5)	48 (60.8)	43 (45.7)	41 (47.7)	43 (60.6)	37 (44.6)
Frontal lobe	8 (8.2)	5 (6.3)	11 (11.7)	7 (8.1)	5 (7.0)	9 (10.8)
Temporal lobe	13 (13.4)	3 (3.8)	3 (3.2)	12 (14.0)	1 (1.4)	3 (3.6)
Parietal lobe	14 (14.4)	7 (8.9)	11 (11.7)	13 (15.1)	6 (8.5)	9 (10.8)
Occipital lobe	5 (5.2)	0 (0.0)	3 (3.2)	4 (4.7)	0 (0.0)	3 (3.6)
Cerebellum	1 (1.0)	0 (0.0)	0 (0.0)	1 (1.2)	0 (0.0)	0 (0.0)
Not known	25 (25.8)	23 (29.1)	33 (35.1)	22 (25.6)	21 (29.6)	31 (37.3)
Lateralisation (if not brain stem), n (%)						
Right side	9 (9.3)	1 (1.3)	5 (5.3)	9 (10.5)	1 (1.4)	5 (6.0)
Left side	79 (81.4)	73 (92.4)	81 (86.2)	69 (80.2)	65 (91.5)	71 (85.5)
Not known	12 (12.4)	5 (6.3)	11 (11.7)	11 (12.8)	5 (7.0)	10 (12.0)
Time post stroke (years)						
	n = 97	n = 79	n = 94	n = 86	n = 71	n = 83
Mean (SD)	2.8 (2.7)	3.4 (4.6)	2.8 (2.9)	2.8 (2.6)	3.6 (4.8)	2.9 (2.9)
Median (IQR)	1.9 (0.9–3.8)	1.9 (1.0–4.3)	1.8 (0.7–3.6)	1.9 (0.9–4.0)	2.1 (1.0–4.5)	1.9 (0.7–3.6)
Min., max.	0.3, 15.7	0.4, 36.1	0.4, 12.7	0.3, 15.7	0.4, 36.1	0.4, 12.7

IQR, interquartile range; max., maximum; min., minimum.

a Based on a CAT Comprehension of Spoken Sentences with possible total scores ranging from 0 to 32.

b Two sites in same NHS trust.

c Based on the CAT Naming Objects with possible total scores ranging from 0 to 48.

Notes

Adapted from Palmer *et al.*⁵⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

For the 270 randomised participants meeting the MI inclusion criteria, the recruitment across 21 speech and language therapy centres/departments (20 NHS trusts) ranged from 7 to 22 participants, with a median of 12 [interquartile range (IQR) 10–15] participants. The majority of participants ($n = 164$; 60.7%) were male. The overall mean age at consent was 65.4 (SD 12.9) years, ranging from 23.1 to 91.8 years. Most participants ($n = 119$; 44.1%) had mild word-finding difficulty; 80 participants (29.6%) had moderate and 71 (26.3%) had severe difficulty. The overall median time post stroke was approximately 2 years (IQR 1–4 years). Most participants showed no evidence of apraxia of speech ($n = 174$; 64.4%) and 212 (78.5%) had suffered an infarction stroke.

In summary, randomised participants and 'completers' were very similar on average with respect to measured characteristics and demographics (see Table 3). Furthermore, on average, the participants appeared broadly similar across interventions. However, by chance, there were a few exceptions indicating relatively small differences between interventions, such as with respect to the location of the stroke, time post stroke, the type of stroke and lateralisation, if not brain stem. There was also a chance imbalance in the number of participants randomised to AC compared with UC or CSLT.

Table 4 shows the characteristics of the randomised participants and 'completers' (with their available carers) with respect to continuous covariates (outcomes assessed at baseline), which are broadly similar between the two MI and mITT populations and across interventions. However, the available carers of

TABLE 4 Baseline characteristics of randomised participants (outcome covariates)

Variable	MI population (N = 270)			mITT population (N = 240)		
	UC (n = 97)	AC (n = 79)	CSLT (n = 94)	UC (n = 86)	AC (n = 71)	CSLT (n = 83)
Word-finding ability (%)^a	n = 97	n = 79	n = 94	n = 86	n = 71	n = 83
Mean (SD)	42.8 (18.1)	41.4 (20.7)	43.2 (19.0)	42.6 (18.1)	41.7 (20.6)	43.7 (19.0)
Median (IQR)	44.0 (30.0–57.0)	37.5 (23.5–59.0)	43.8 (30.0–57.5)	42.3 (30.0–57.0)	37.5 (25.0–59.0)	43.0 (30.0–58.2)
Min., max.	5.0, 85.0	9.0, 82.0	4.5, 86.0	5.0, 85.0	9.5, 82.0	4.5, 86.0
Functional communication (TOMs)^b	n = 96	n = 78	n = 93	n = 86	n = 70	n = 82
Mean (SD)	3.1 (1.0)	2.7 (1.0)	2.9 (1.2)	3.1 (1.0)	2.7 (1.1)	2.9 (1.2)
Median (IQR)	3.0 (2.5–4.0)	2.5 (2.0–3.5)	3.0 (2.0–4.0)	3.0 (2.5–4.0)	2.5 (2.0–3.5)	3.0 (2.0–4.0)
Min., max.	0.5, 5.0	1.0, 4.5	0.5, 5.0	0.5, 5.0	1.0, 4.5	0.5, 5.0
COAST (%)^c	n = 94	n = 79	n = 89	n = 84	n = 71	n = 79
Mean (SD)	59.9 (13.1)	60.0 (13.8)	58.2 (13.6)	59.8 (13.2)	59.5 (14.0)	58.4 (13.6)
Median (IQR)	61.3 (52.5–68.8)	60.0 (48.8–68.8)	57.5 (48.8–68.8)	61.3 (51.9–68.8)	60.0 (48.8–67.5)	57.5 (47.5–68.8)
Min., max.	26.3, 86.3	26.3, 96.3	26.3, 87.5	26.3, 86.3	26.3, 96.3	26.3, 87.5
CaCOAST 15 (%)^d	n = 58	n = 49	n = 62	n = 53	n = 44	n = 56
Mean (SD)	56.8 (14.9)	53.7 (13.2)	52.8 (15.6)	56.5 (14.7)	54.0 (13.3)	53.6 (14.7)
Median (IQR)	57.5 (46.7–66.7)	51.7 (43.3–66.7)	50.8 (41.7–63.3)	58.3 (46.7–66.7)	52.5 (43.3–66.7)	50.8 (42.5–64.2)
Min., max.	26.7, 81.7	28.3, 78.3	18.3, 81.7	26.7, 81.7	28.3, 78.3	20.0, 81.7
CaCOAST 5 (%)^e	n = 58	n = 49	n = 62	n = 53	n = 44	n = 56
Mean (SD)	54.7 (19.3)	44.7 (16.4)	48.2 (21.0)	54.1 (18.5)	45.1 (16.7)	48.7 (20.5)
Median (IQR)	55.0 (40.0–70.0)	50.0 (30.0–55.0)	47.5 (30.0–65.0)	55.0 (40.0–65.0)	50.0 (30.0–55.0)	50.0 (32.5–65.0)
Min., max.	20.0, 100.0	10.0, 95.0	5.0, 90.0	20.0, 100.0	10.0, 95.0	5.0, 90.0

Variable	MI population (N = 270)			mITT population (N = 240)		
	UC (n = 97)	AC (n = 79)	CSLT (n = 94)	UC (n = 86)	AC (n = 71)	CSLT (n = 83)
Word-finding of untreated words^f (CAT score)	n = 97	n = 79	n = 94	n = 86	n = 71	n = 83
Mean (SD)	26.4 (11.0)	26.2 (11.5)	26.5 (11.4)	26.2 (11.0)	26.6 (11.3)	26.6 (11.3)
Median (IQR)	28.0 (18.0–36.0)	30.0 (16.0–36.0)	27.5 (17.0–38.0)	27.5 (17.0–35.0)	30.0 (16.0–37.0)	27.0 (17.0–38.0)
Min., max.	5.0, 43.0	6.0, 42.0	5.0, 43.0	5.0, 43.0	6.0, 42.0	5.0, 43.0
EQ-5D-5L VAS score (patients - aphasia friendly)^g	n = 97	n = 79	n = 94	n = 86	n = 71	n = 83
Mean (SD)	69.8 (17.6)	68.9 (20.0)	67.0 (21.1)	70.0 (17.5)	68.7 (20.3)	67.0 (21.5)
Median (IQR)	75.0 (55.0–80.0)	75.0 (55.0–85.0)	70.0 (50.0–85.0)	75.0 (60.0–80.0)	70.0 (55.0–85.0)	70.0 (50.0–85.0)
Min., max.	30.0, 100.0	10.0, 100.0	10.0, 100.0	30.0, 100.0	10.0, 100.0	10.0, 100.0
EQ-5D-5L index (patients - aphasia friendly)^h	n = 97	n = 79	n = 93	n = 86	n = 71	n = 83
Mean (SD)	0.72 (0.20)	0.70 (0.22)	0.70 (0.22)	0.75 (0.17)	0.70 (0.22)	0.70 (0.22)
Median (IQR)	0.75 (0.61–0.88)	0.75 (0.57–0.87)	0.75 (0.61–0.87)	0.76 (0.66–0.88)	0.75 (0.57–0.89)	0.75 (0.61–0.87)
Min., max.	0.21, 1.00	0.05, 1.00	0.02, 1.00	0.21, 1.00	0.05, 1.00	0.02, 1.00
EQ-5D-5L index (carer)^h	n = 59	n = 49	n = 63	n = 53	n = 44	n = 57
Mean (SD)	0.85 (0.18)	0.83 (0.18)	0.82 (0.19)	0.85 (0.18)	0.83 (0.19)	0.82 (0.19)
Median (IQR)	0.90 (0.82–0.95)	0.87 (0.78–0.92)	0.87 (0.75–0.92)	0.90 (0.82–0.95)	0.88 (0.78–0.92)	0.87 (0.75–0.92)
Min., max.	0.21, 1.00	0.19, 1.00	-0.00, 1.00	0.21, 1.00	0.19, 1.00	-0.00, 1.00
EQ-5D-5L VAS score (carer)^g	n = 58	n = 49	n = 63	n = 53	n = 44	n = 57
Mean (SD)	79.5 (15.0)	76.6 (19.0)	76.0 (18.8)	79.2 (15.3)	77.7 (17.7)	75.4 (19.2)
Median (IQR)	80.0 (75.0–90.0)	80.0 (65.0–90.0)	80.0 (65.0–90.0)	80.0 (75.0–90.0)	80.0 (70.0–90.0)	80.0 (65.0–90.0)
Min., max.	25.0, 100.0	20.0, 98.0	25.0, 100.0	25.0, 100.0	20.0, 98.0	25.0, 100.0

continued

TABLE 4 Baseline characteristics of randomised participants (outcome covariates) (continued)

Variable	MI population (N = 270)			mITT population (N = 240)		
	UC (n = 97)	AC (n = 79)	CSLT (n = 94)	UC (n = 86)	AC (n = 71)	CSLT (n = 83)
EQ-5D-5L VAS score (proxy)^g	n = 73	n = 56	n = 65	n = 64	n = 49	n = 59
Mean (SD)	62.5 (18.9)	64.1 (21.9)	62.0 (21.6)	62.6 (18.9)	65.4 (21.1)	62.0 (21.6)
Median (IQR)	65.0 (50.0–80.0)	70.0 (50.0–80.0)	60.0 (45.0–80.0)	65.0 (50.0–80.0)	70.0 (55.0–80.0)	60.0 (45.0–80.0)
Min., max.	15.0, 95.0	4.0, 95.0	10.0, 100.0	15.0, 95.0	4.0, 95.0	10.0, 100.0
EQ-5D-5L index (patient – proxy)^h	n = 73	n = 56	n = 64	n = 64	n = 49	n = 58
Mean (SD)	0.63 (0.23)	0.64 (0.21)	0.61 (0.24)	0.64 (0.23)	0.65 (0.22)	0.60 (0.24)
Median (IQR)	0.68 (0.49–0.78)	0.70 (0.51–0.77)	0.65 (0.39–0.79)	0.69 (0.51–0.78)	0.71 (0.51–0.78)	0.63 (0.39–0.79)
Min., max.	-0.11, 1.00	-0.06, 1.00	0.04, 1.00	-0.11, 1.00	-0.06, 1.00	0.04, 1.00

IQR, interquartile range; max., maximum; min., minimum.

a Word-finding ability of personally selected words (%) based on the personal vocabulary naming test.

b TOMs rating score ranges from 0 to 5, with higher scores meaning improved functional communication.

c Higher score indicates positive self-perceived communication and impact on patient's quality of life.

d Based on the first 15 items of the CaCOAST, with higher scores indicating positive carer perception of patient's communication ability.

e Based on the last five items of the CaCOAST, with higher scores indicating positive carer perception of the impact of the patient's communication ability on the carer's quality of life.

f Based on a CAT Naming Objects, with total scores ranging from 0 to 48; higher scores indicate improved word-finding ability of untreated words.

g Higher score indicates positive perception of health status, with 0 and 100 meaning worst and best health status imaginable.

h Higher values indicate higher health-related quality of life.

Notes

-0.00 means < 0.

Adapted from Palmer *et al.*⁵⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

the UC participants who agreed to take part had slightly higher CaCOAST domain scores on average than their counterparts did. It should also be noted that the participant, not the supporting carer, was the unit of randomisation.

Reported cases of unblinding of outcome assessments

Table 5 summarises cases of suspected unblinding and when the allocated intervention was guessed correctly. At the 6-month assessment, 29 (10.7%) cases of suspected unblinding were reported. Of these, the allocated intervention was incorrectly guessed in only one case. The majority of cases happened before or during the 6-month assessment, and the proportions of participants were higher in the CSLT and AC groups than in the UC group. Nevertheless, the reported cases of unblinding are negligible relative to the number of outcome assessments made during the trial.

Usual-care speech and language therapy offered (fidelity/adherence to provision of usual care)

The objective here was to assess whether or not the overall amounts of UC speech and language therapy were similar across the intervention groups throughout the trial as it was an adjunct trial aiming to compare the interventions in addition to UC with UC alone. The descriptive details of what UC speech and language therapy consisted of are reported in Chapter 2, Usual-care control group. Table 6 shows the proportion of participants who received UC speech and language therapy 3 months prior to each assessment time point. The distribution of UC speech and language therapy seems comparable across interventions; however, slightly fewer participants in the AC group than in the UC and CLST groups received UC speech and language therapy especially 3 months prior to baseline. In addition, slightly more participants in the UC group received UC speech and language therapy than their counterparts. Across interventions, the proportions of participants receiving UC speech and language therapy decreased as the trial progressed. This was to be expected as participants were further in time post stroke as the trial progressed and the pre-baseline UC descriptions (see Chapter 2, Usual-care control group) identified the fact that amounts of speech and language therapy received decreased with length of time post stroke.

The distributions of the average amounts of overall UC speech and language therapy received are displayed in Figure 3; the averages are calculated based on all participants in the denominator regardless of whether or not they received UC speech and language therapy. On average, during the 6-month intervention period (data collection points 3 and 6 months), the UC group received a mean of 3.8 hours of usual speech and language therapy, the AC group received a mean of 3.2 hours of usual speech and language therapy and the CSLT group received a mean of 3.2 hours of usual speech and language therapy. Those in receipt of therapy had a mean of 9.7 hours in UC, 11 hours in AC and 7.8 hours in CSLT. There is more variation in the average number of hours in receipt of therapy in those who received therapy than for all participants in each group as the denominator of those in receipt was small, as shown in Table 6.

TABLE 5 Cases of unblinding of outcome assessments

Classification	Assessment	Trial group, n (%)		
		UC (N = 97)	AC (N = 79)	CSLT (N = 94)
Suspected unblinding	6 months	3 (3.1)	9 (11.3)	17 (18.1)
	9 months	2 (2.1)	4 (5.0)	3 (3.2)
	12 months	0 (0.0)	1 (1.3)	1 (1.1)
Unblinding ^a	6 months	2 (2.1)	8 (10.0)	17 (18.1)
	9 months	1 (1.0)	4 (5.0)	3 (3.2)
	12 months	0 (0.0)	0 (0.0)	1 (1.1)

a Correctly guessed the intervention.

TABLE 6 Receivers of overall UC speech and language therapy

Time point	Trial group, n (%)		
	UC (N = 96) ^a	AC (N = 79)	CSLT (N = 94)
Baseline	43 (44.8)	30 (38.0)	42 (44.7)
3 months	33 (34.4)	22 (27.8)	30 (31.9)
6 months	23 (24.0)	14 (17.7)	21 (22.3)
9 months	17 (17.7)	12 (15.2)	12 (12.8)
12 months	13 (13.5)	10 (12.7)	8 (8.5)

a UC speech and language therapy data were not collected for one participant.

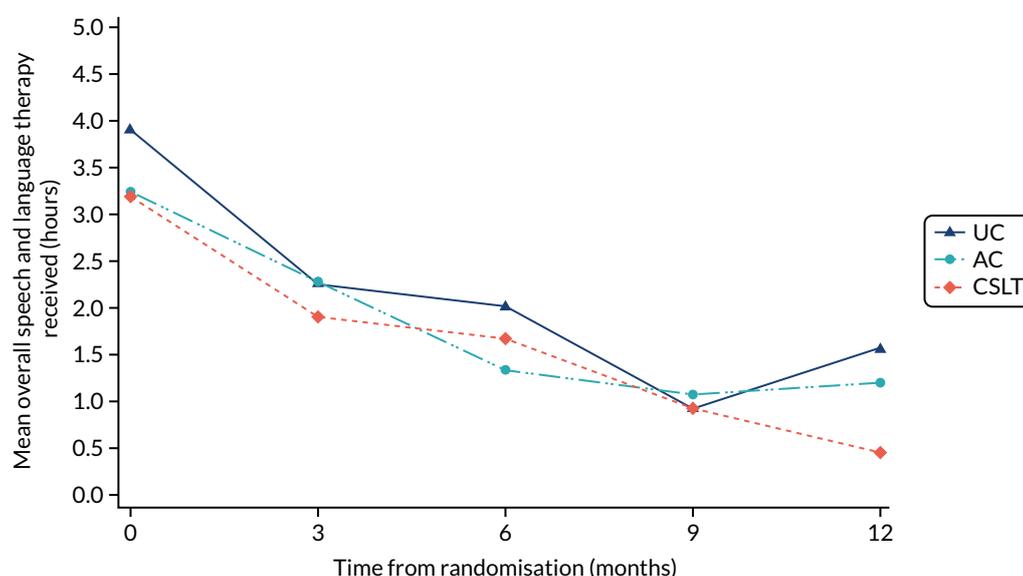


FIGURE 3 Mean amounts of overall UC speech and language therapy (all participants).

The mean amounts of overall UC speech and language therapy for only those who received it can be seen in figure 6 of the full statistical report, which is available online as a supplementary appendix to Palmer *et al.*⁵⁰

Fidelity to computerised speech and language therapy: adherence to practice and quality of intervention delivery

Time to access computer therapy

All participants randomised to the CSLT group were given access to the computer software (full intervention coverage was achieved). The median time to access computer therapy since randomisation among 94 CSLT participants was approximately 26 days (IQR 19–35 days), with a range of 7 to 114 days. This excludes three deaths prior to the 6-month assessment. Most participants accessed computer therapy within 1 month from randomisation. There were two outliers who accessed computer therapy after 3.5 months (112 and 114 days) from randomisation. In the first case, there were several reasons for the delay: there were issues with the participant’s computer and StepByStep software version on the Big CACTUS laptop, the therapist went on sick leave and then study leave, and the participant was also hospitalised with cardiac problems. In the second case, the SLT delivering the intervention went on long-term sick leave and later resigned, so there was no one available to set the computer up for the participant.

Computer therapy practice time

Table 7 summarises the distributions of the total computer practice time and the mean computer practice time (per month and per week) within 6 months from randomisation with the number of computer therapy sessions contributing to the practice time. The mean total computer practice time per participant over 6 months was 28 hours and the median was 21.1 hours, with an IQR of 4.9–49.7 hours and a range of 0–104.5 hours. Appendix 5, Figure 26, shows the patterns of computer practice time per participant throughout the trial including continued computer use after 6 months. Of the 94 CSLT participants, 57 (60.6%) continued to use the computer therapy beyond the 6 months of randomisation.

Per-protocol adherence to computerised speech and language therapy

Three of the 97 participants randomised to CSLT died prior to the 6-month assessment. Of the remaining 94, 60 (63.8%) and 43 (45.7%) used the computer therapy for at least 10 hours and 26 hours within 6 months of randomisation, respectively. These participants were deemed to have adhered to the CSLT intervention in terms of practice time. In addition, no computer use at all was recorded for 11 participants (11.7%) within 6 months of randomisation. Six of these participants withdrew from the intervention owing to illness ($n = 2$), difficulty using the computer ($n = 3$) or time commitment of the intervention ($n = 1$). In the remaining cases no data were recorded because key files that contain practice data were corrupted or were not collected from computers. In such cases, it is not known whether practice was conducted or not. However, average practice times in Table 7 are based on 0 hours for those with no key file data, potentially underestimating the average practice time.

Quality of computerised speech and language therapy intervention delivery

Therapist knowledge about the computerised speech and language therapy intervention

At the 5-month quiz, therapists' median score was 10 (range 7–13) out of a possible 15, indicating that therapists generally had a reasonable level of knowledge about the intervention. The level of knowledge about the intervention increased over time as therapists delivered the intervention to more participants; the median score was 12 at the 10-month quiz, which was maintained at the 15-month quiz.

Tailoring of the computerised speech and language therapy intervention

Based on the therapy planning form ratings, the StepByStep approach expert had a comprehensive understanding of why the therapist had tailored the steps as they had (score of 2) for 66% of therapy planning forms, had some understanding of why the steps had been tailored in this way (score of 1) for 24% and was not clear why the steps had been tailored as they had (score of 0) for 10% of the sample. This suggests that the majority of SLTs tailored the therapy as intended by the author of the therapy manual following the speech and language therapy principles for treating a word-finding impairment.

TABLE 7 Summary of the distribution of computer practice time ($n = 94$)

Computer use classification	Mean (SD)	Median (IQR)	Min., max.
Computer practice time within 6 months			
Total (hours)	28.0 (25.6)	21.1 (4.9–49.7)	0.0, 104.5
Average per month (hours)	4.7 (4.3)	3.5 (0.8–8.3)	0.0, 17.4
Average per week (minutes)	64 (59)	49 (11–114)	0, 240
Number of computer sessions			
Total	60 (49)	58 (14–100)	0, 177
Average per week	2.3 (1.9)	2.2 (0.5–3.8)	0.0, 6.8
Max., maximum; min., minimum.			

Speech and language therapist activities in delivering the computerised speech and language therapy intervention

The median amount of time spent setting up and supporting the participant over the intervention period was 5 hours and 55 minutes (ranging from 30 minutes to 26 hours) over a median of four sessions (range 1–22 sessions). This can be further broken down into the median amount of time spent interacting with the participants, including face-to-face, telephone or e-mail contact, of 1 hour and 45 minutes (range 0 minutes to 11 hours) over three sessions (range 0–19 sessions) and the median amount of time the therapist spent alone setting up the StepByStep software of 4 hours (range 0 minutes to 18 hours and 10 minutes) over one session (range 0–11 sessions). Although a greater number of the sessions were spent interacting with the participant, the therapists generally spent more time alone tailoring the StepByStep software before providing it to the participants for word-finding practice.

As well as delivering the intervention to the participants, SLTs also trained and supported the volunteer or SLTA. The median time SLTs spent with each participant's volunteer or assistant was 1 hour and 40 minutes (range 20 minutes to 8 hours and 35 minutes) across a median of four sessions (range 1–13 sessions). The majority of assistants/volunteers supporting participants were provided with training (93%) and ongoing support (84%) from the therapists. Training was typically delivered once for 1 hour, whereas the median number of support sessions was two, lasting a median of 30 minutes. Approximately half (52%) of the SLTAs or volunteers supporting the participants had their feedback forms monitored by the SLT. The therapy manual states that the assistant or volunteer should complete a feedback form and return it to the SLT each time they see the participant and 'the SLT should use this to monitor the volunteer/assistant support and the progress of the patient'. This suggests that the two-way communication between the SLT and the volunteer/assistant was not conducted for half of the participants. For the SLTAs and volunteers whose feedback forms were monitored, the monitoring took place a median of three times, taking 20 minutes in total.

Volunteer/speech and language therapy assistant activities in delivering the computerised speech and language therapy intervention

The volunteer/SLTA activity logs were completed for 86 of the 97 CSLT intervention group participants. One participant was recorded to have declined all support from a SLTA/volunteer. Either the other 10 participants did not receive support, suggesting partial coverage of this active ingredient of the intervention, or the activity logs were not completed. Of the 86 participants with completed activity logs, a median number of five sessions (range 1–12 sessions) took place between the volunteer/assistant and the participant over a median of 4 hours and 15 minutes (range 20 minutes to 8 hours and 45 minutes). The actual amount of input delivered by the volunteer/assistant was lower than the therapy manual's recommendation of 6 hours of input over 6 months, but equal to the minimum amount defined as per protocol accounting for periods of holiday and illness. The vast majority of participants for whom an activity log was completed received encouragement and motivation to use the computer therapy from the volunteer/assistant (99%), which was delivered for a median of 1 hour and 25 minutes across four sessions. Other activities were carried out with fewer participants. Although this might have been due to lack of need in the case of setting up or adjusting the computer or microphone (87%) or assistance with using the software (90%), it is likely that encouraging the use of new words through practising them in conversation has the potential to be useful for all participants, but it was carried out with only 85% of participants for a median of 45 minutes (range 5 minutes to 2 hours and 35 minutes).

Per-protocol adherence to the attention control intervention

Only 1 of the 80 participants randomised to the AC intervention died before the 6-month assessment. Of the remaining 79, only 14 (17.7%) were sent at least six puzzle books and were contacted at least four times within 6 months. These were deemed to have adhered to the predefined key components of the AC group as detailed in *Chapter 3, Per-protocol sets*.

Using four puzzle books and four contacts as per protocol in the post hoc analysis to be consistent with the 4 months of computer therapy as per protocol in the CSLT group, 48 out of 79 (60.8%) participants were deemed to have adhered to the AC intervention. This post hoc approach was used only for sensitivity analysis (see *Chapter 3, Per-protocol sets*).

Response profiles of participants

Figure 4 shows the changes in the response of participants with respect to word-finding ability over time, stratified by intervention. There appear to be marked improvements in the CSLT group compared with the UC or AC groups. Only 10 CSLT participants failed to improve their word-finding ability at 6 months; three had no computer use recorded, four used computer therapy for < 5 hours in the first 6 months and the remaining three used computer therapy for > 20 hours in the first 6 months.

Figure 5 shows the pattern in functional communication responses, which is unclear but appears to be similar across interventions. Only five (1.9%) participants had a TOMs rating on the ceiling (score of 5): 3 out of 94 (3.2%) in the CSLT group, 2 out of 97 (2.1%) in the UC group and none in the AC group. These participants cannot show any further improvements in functional communication during the trial because they were deemed to communicate effectively in all situations at baseline. It should be noted that the ability to communicate effectively in all situation does not preclude having word-finding difficulties as communication can be achieved by other means, such as gesture. No participants were unable to communicate in any way (with a TOMs score of 0) at baseline.

The COAST response patterns over time, indicating participant perception of social participation and quality of life, stratified by intervention, are displayed in Figure 6. Participants in the AC group appear to have deteriorated at the 6-month assessment relative to their baseline. These response profile plots complement the interpretation of the main and subgroup results presented in *Chapter 5, Results*.

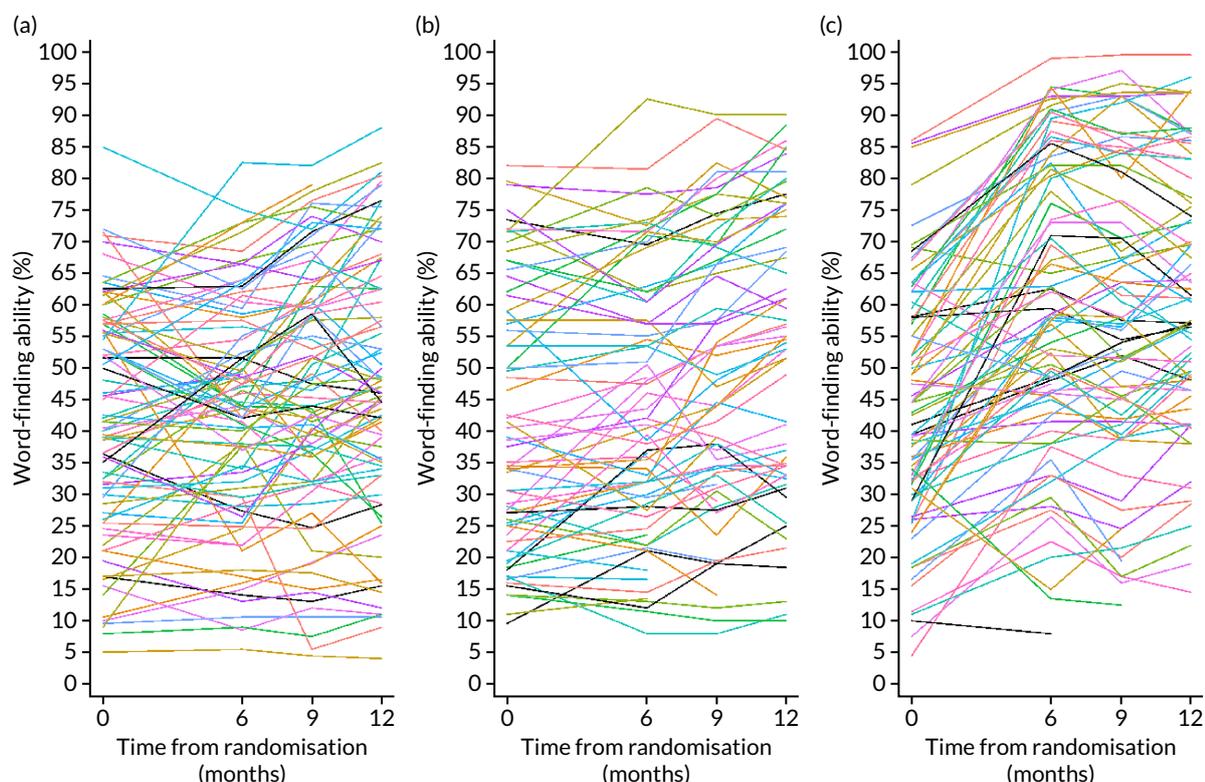


FIGURE 4 Changes in word-finding over time stratified by the intervention. (a) UC; (b) AC; and (c) CSLT. Each line, regardless of colour, indicates a participant's response profile.

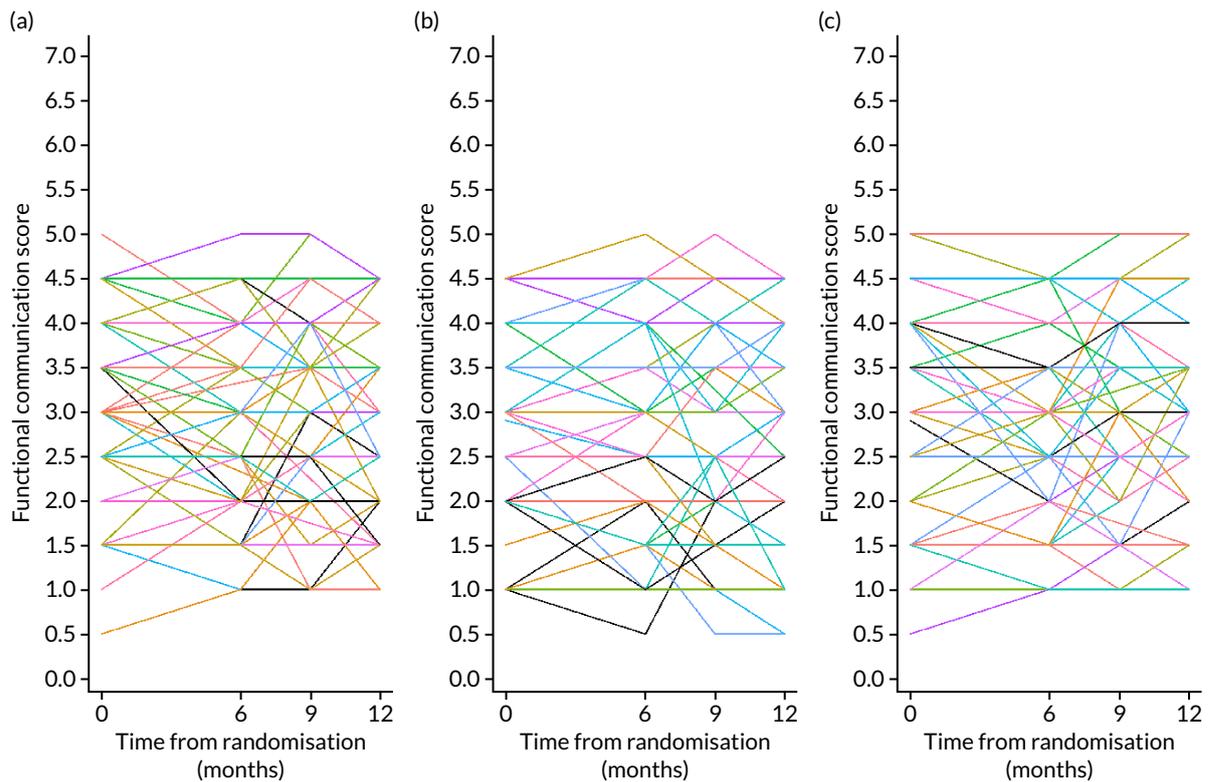


FIGURE 5 Changes in functional communication in conversation rated by TOMs over time, stratified by intervention. (a) UC; (b) AC; and (c) CSLT. Each line, regardless of colour, indicates participant's response profile.

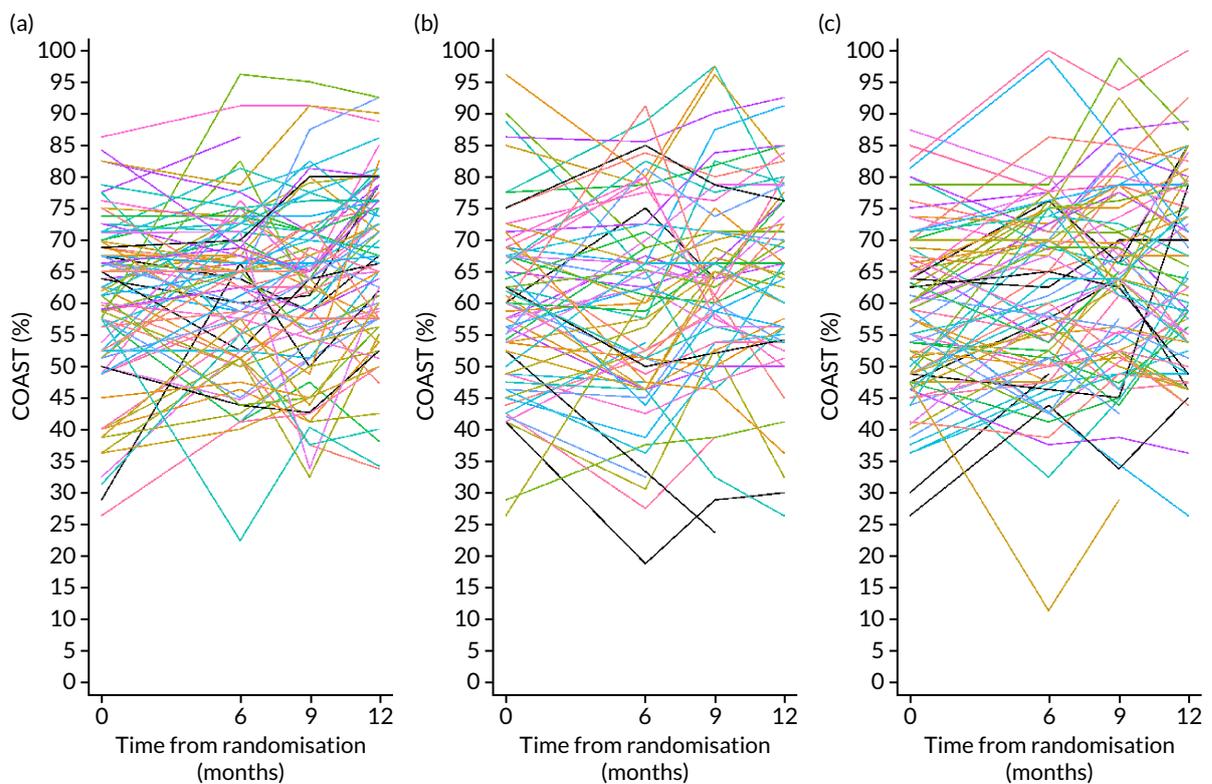


FIGURE 6 Changes in perception of social participation and quality of life rated by COAST over time, stratified by intervention. (a) UC; (b) AC; and (c) CSLT. Each line, regardless of colour, indicates a participant's response profile.

Effect of the intervention on the co-primary and key secondary outcome measures

The first three primary clinical effectiveness objectives were to establish whether or not self-managed CSLT intervention increases the ability of people with aphasia to use the vocabulary of personal importance, improves functional communication ability in conversation and results in perceived greater changes in social participation in daily activities and quality of life. This section presents results at the 6-month assessment to address these objectives.

Word-finding and functional communication: co-primary outcome measures

Figures 7 and 8 display the unadjusted mean responses to word-finding and functional communication over time stratified by the intervention.

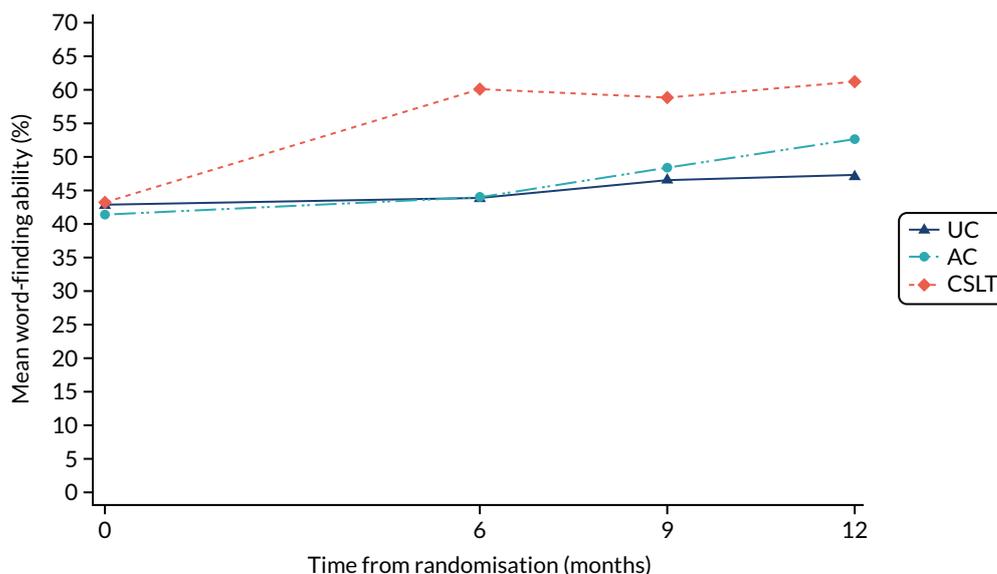


FIGURE 7 Mean word-finding ability over time stratified by the intervention. Adapted from Palmer *et al.*⁵⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

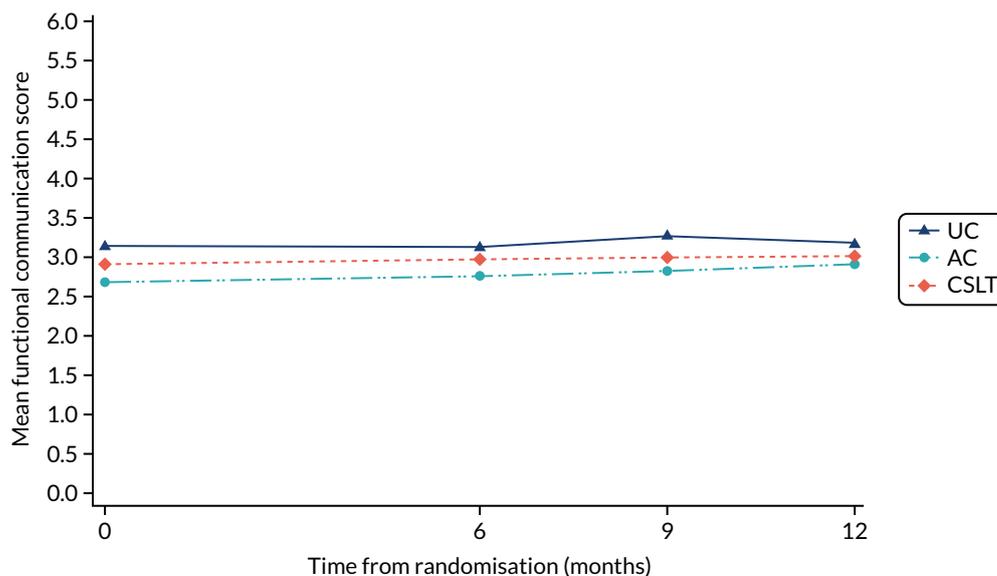


FIGURE 8 Mean functional communication in conversation over time stratified by the intervention. Adapted from Palmer *et al.*⁵⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

For word-finding, 86, 71 and 83 participants in the UC, AC and CSLT groups, respectively, were included in the mITT analysis. The mean improvement in word-finding of personally selected words was 1.1% (SD 11.2%) in the UC group, 2.4% (SD 8.8%) in the AC group and 16.4% (SD 15.3%) in the CSLT group. This indicates an adjusted mean difference in word-finding improvement of 16.2% (95% CI 12.7% to 19.6%; $p < 0.0001$) in favour of CSLT compared with UC. Detailed results are presented in *Table 8*.

In *Figure 8*, the mean functional communication barely changed at 6 months. As shown in *Table 8*, the mean change in functional communication in conversation on the TOMs was very similar between the CSLT and UC groups, translating to an adjusted MDC of -0.03 (95% CI -0.21 to 0.14 ; $p = 0.709$) compared with UC.

In line with the prespecified Hochberg multiple testing strategy, shown in *Figure 1*, we can only claim the clinical effectiveness of CSLT compared with UC in improving word-finding of personal choice at a 2.5% significance level. Because both comparisons between CSLT and UC with respect to word-finding and functional communication in conversation were not statistically significant at a 5% significance level, further statistical significance testing is prohibited. The mean improvement in word-finding of personally selected words of 14.4% (95% CI 10.8% to 18.1%) in favour of CSLT compared with AC supports that the clinical effectiveness in improving word-finding is attributed to CSLT rather than the attention provided. The mean improvement in word-finding between the AC and UC groups was similar: 1.8% (95% CI -1.9% to 5.4%). The mean changes in functional communication in conversation were very similar across interventions. Thus, CSLT did not result in improvement in functional communication ability in conversation compared with UC or AC.

The effects of CSLT on word-finding and functional communication were very similar after adjusting for additional prespecified covariates (time post stroke and location of stroke) (*Tables 8 and 9*).

Patient perception of communication and its impact on their life: key secondary outcome measure

Figure 9 shows the unadjusted average profile in COAST over time. The effect of the intervention on the COAST at 6 months is presented in *Table 8*.

The average improvement in COAST was only 3.3% and 2.7% in the CSLT and UC groups, respectively, and the AC group COAST decreased slightly, by 0.3%. The adjusted MDC in COAST was only 0.5% (95% CI -3.1% to 4.1%) in the CSLT group compared with the UC group. The adjusted MDC of 3.8% (95% CI -0.0% to 7.5%) observed in the CSLT group compared with the AC group was attributable to the AC group barely changing on average at 6 months, whereas the UC group experienced a small average improvement that was comparable to that seen in the CSLT group.

In summary, there is insufficient evidence to show that CSLT improves patients' perceptions of communication and its impact on their life.

Impact of attrition and adherence on the co-primary and key secondary outcome measures

As part of the sensitivity analysis, we present the results exploring the influence of attrition and intervention adherence on the effect of the intervention on word-finding, functional communication in conversation (TOMs) and participant perception of social participation and quality of life (COAST) at 6 months. The corresponding results for the MI set, four per-protocol sets and linear interpolation set described in *Chapter 3, Analysis populations*, are presented in forest plots (*Figures 10–12*), including the primary mITT results presented in *Table 8* for comparability.

In general, the results are consistent across all analysis sets considered and similar to the primary analysis set (mITT). The results of per-protocol comparisons that involve the AC group should be interpreted with caution because of small sample size due to poor adherence to AC of 17.7%, based on the large number

TABLE 8 Co-primary and key secondary outcomes at 6 months (mITT)

Co-primary and key secondary outcomes at 6 months	UC		AC		CSLT		CSLT vs. UC ^g		CSLT vs. AC ^h		AC vs. UC ^g	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Adjusted MDC (95% CI)	p-value	Adjusted MDC (95% CI)	p-value	Adjusted MDC (95% CI)	p-value
Co-primary outcomes												
Change in word-finding (%) ⁱ	86	1.1 (11.2)	71	2.4 (8.8)	83	16.4 (15.3)	16.2 (12.7 to 19.6) ^a	< 0.0001	14.4 (10.8 to 18.1) ^c	< 0.0001	1.8 (-1.9 to 5.4)	0.338
Change in functional communication ^j	84	0.05 (0.59)	68	0.10 (0.61)	81	0.04 (0.58)	-0.03 (-0.21 to 0.14) ^b	0.709	-0.01 (-0.20 to 0.18) ^d	0.915	-0.02 (-0.21 to 0.17)	0.812
Key secondary outcome												
Change in COAST (%) ^k	83	2.7 (12.6)	68	-0.3 (12.7)	82	3.3 (11.3)	0.5 (-3.1 to 4.1) ^e	0.772	3.8 (-0.0 to 7.5) ^f	0.051	-3.2 (-7.0 to 0.5)	0.089

Superscripts a, b, c, d, e and f are referenced in *Figure 1* to aid interpretation of Hochberg sequential and hierarchical hypotheses testing procedure for claiming evidence.

g UC as the reference group.

h AC as the reference group.

i Higher scores indicate improved vocabulary of personal importance.

j Higher scores indicate improved functional communication ability in conversation. Seven participants had missing TOMs data (video not recorded in error, poor video sound quality, camera technical issues, participant declined to be videoed, recording failed owing to a technical issue, participant was unwell and did not want to complete the assessment and participant did not want to complete assessment).

k Higher percentage score indicates improved patient perception of communication effectiveness and its impact on their quality of life. Seven participants had invalid COAST records, with > 10% of applicable items that were unclear or had no response.

Notes

Results are based on a multiple linear regression model adjusted for baseline measures and fixed stratification factors (centre and severity of word-finding difficulty).

The comparisons of interest are CSLT vs. UC and CSLT vs. AC. AC vs. UC is presented for completeness but shaded to indicate that it is not of clinical importance.

Adapted from Palmer *et al.*⁵⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

TABLE 9 Sensitivity analysis of the co-primary and key secondary outcomes at 6 months (mITT)

Sensitivity analysis: co-primary and key secondary outcomes at 6 months	UC		AC		CSLT		CSLT vs. UC ^a		CSLT vs. AC ^b		AC vs. UC ^c	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Adjusted MDC (95% CI)	p-value	Adjusted MDC (95% CI)	p-value	Adjusted MDC (95% CI)	p-value
Co-primary outcomes												
Change in word-finding (%) ^c	86	1.1 (11.2)	71	2.4 (8.8)	83	16.4 (15.3)	16.3 (12.8 to 19.8)	< 0.0001	14.7 (11.0 to 18.4)	< 0.0001	1.6 (-2.1 to 5.4)	0.385
Change in functional communication ^d	84	0.05 (0.59)	68	0.10 (0.61)	81	0.04 (0.58)	-0.05 (-0.23 to 0.13)	0.596	-0.03 (-0.22 to 0.16)	0.781	-0.02 (-0.22 to 0.17)	0.830
Key secondary outcome												
Change in COAST (%) ^e	83	2.7 (12.6)	68	-0.3 (12.7)	82	3.3 (11.3)	0.9 (-2.8 to 4.5)	0.644	3.6 (-0.2 to 7.5)	0.064	-2.8 (-6.6 to 1.1)	0.156

a UC as the reference group.

b AC as the reference group.

c Higher scores indicate improved vocabulary of personal importance.

d Higher scores indicate improved functional communication ability in conversation. Seven participants had missing TOMs data (video not recorded in error, poor video sound quality, camera technical issues, participant declined to be videoed, recording failed owing to a technical issue, participant was unwell and did not want to complete the assessment and participant did not want to complete assessment).

e Higher percentage score indicates improved patient perception of communication effectiveness and its impact on their quality of life. Seven participants had invalid COAST records, with > 10% of applicable items that were unclear or had no response.

Notes

Sensitivity analysis results from multiple linear regression models adjusted for baseline measures, fixed stratification factors (centre and severity of word-finding) and potential confounders (length of time post stroke and location of stroke).

The comparisons of interest are CSLT vs. UC and CSLT vs. AC. AC vs. UC is presented for completeness but shaded to indicate that it is not of clinical importance.

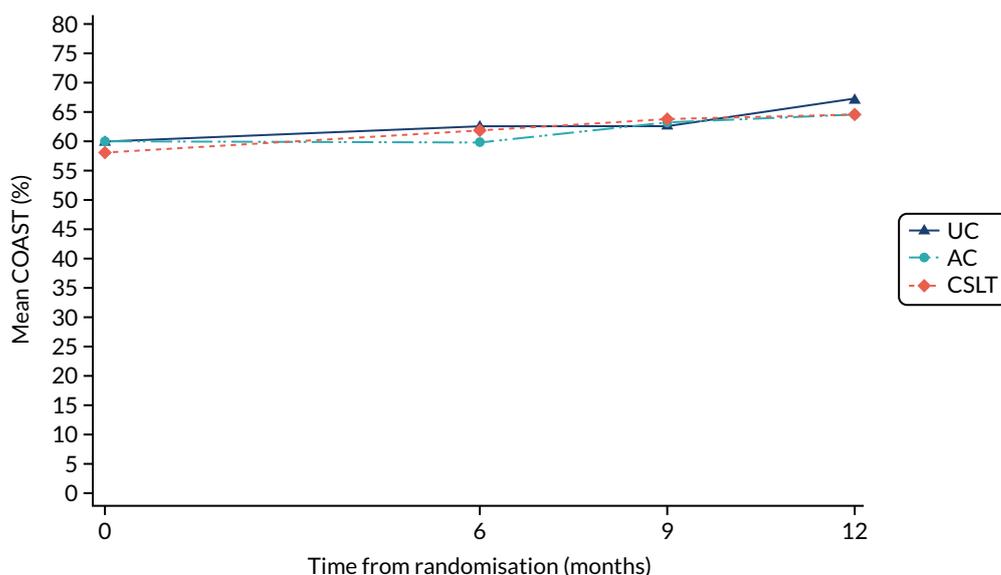


FIGURE 9 Mean COAST score over time, stratified by the intervention. Adapted from Palmer *et al.*⁵⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

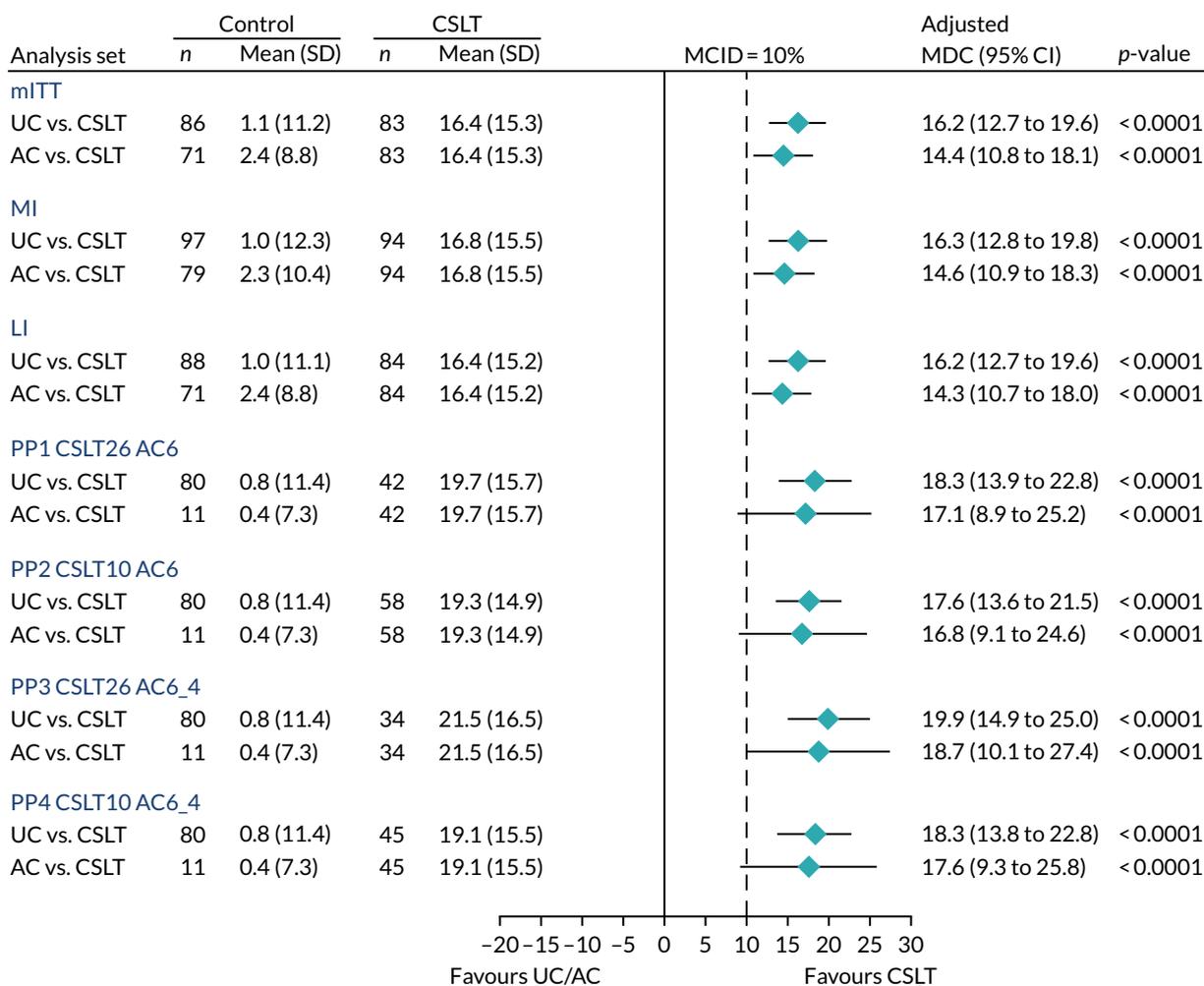


FIGURE 10 Impact of attrition and intervention adherence on word-finding at 6 months. LI, linear interpolation; MCID, minimum clinically important difference; PP, per protocol. Per-protocol sets are defined in *Chapter 3, Analysis populations*.

CLINICAL EFFECTIVENESS RESULTS

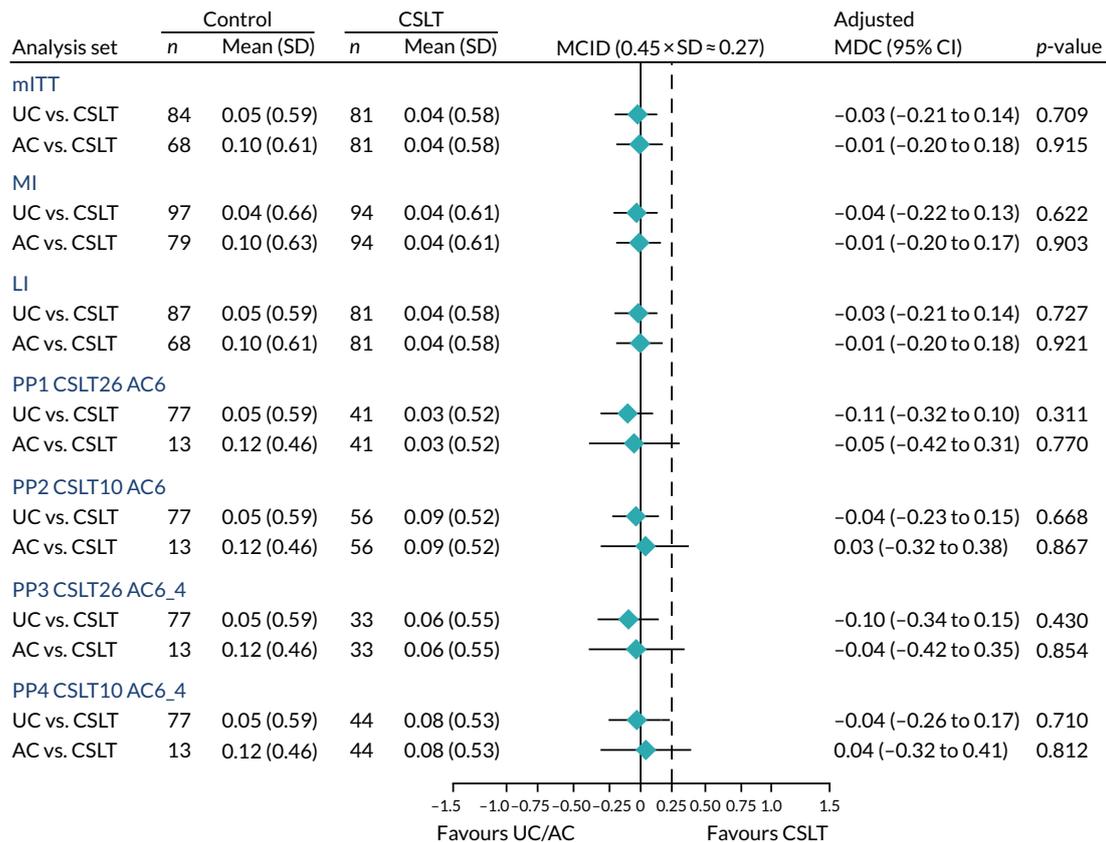


FIGURE 11 Impact of attrition and intervention adherence on functional communication at 6 months. LI, linear interpolation; MCID, minimum clinically important difference; PP, per protocol. Per-protocol sets are defined in Chapter 3, Analysis populations.

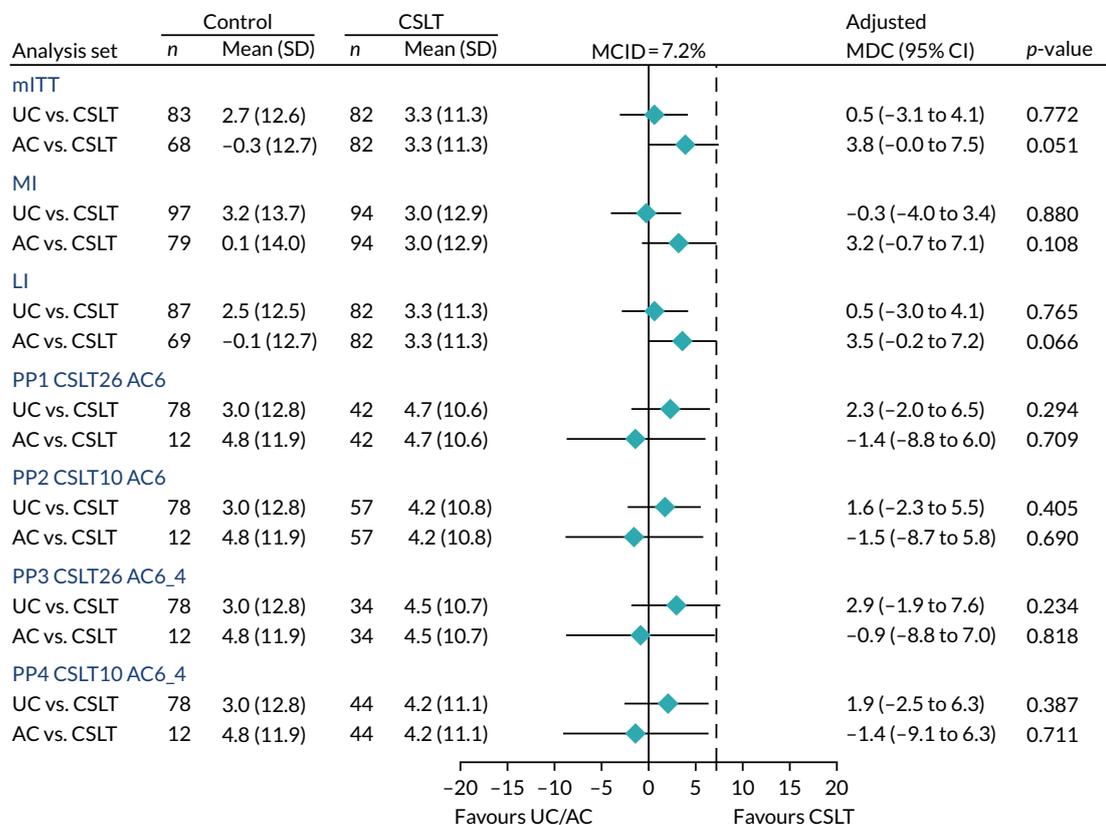


FIGURE 12 Impact of attrition and intervention adherence on COAST at 6 months. LI, linear interpolation; MCID, minimum clinically important difference; PP, per protocol. Per-protocol sets are defined in Chapter 3, Analysis populations.

of puzzle books prespecified as per protocol (six books), as described in *Chapter 3, Per-protocol sets*. We further undertook sensitivity analysis based on post hoc classification of per-protocol adherence in the AC group (sent at least four puzzle books rather than six), as described in *Chapter 3, Per-protocol sets*. These results did not differ from the prespecified per-protocol classification to change the interpretation of findings. For additional information, these post hoc results are accessible online (see section 9.19 of the statistical report, which is available online as a supplementary appendix to Palmer *et al.*⁵⁰).

Subgroup influence on the effectiveness on the co-primary and key secondary outcome measures

This section presents the results exploring potential heterogeneity in the intervention effect across prespecified subgroups on word-finding, functional communication in conversation (TOMs) and participant perception of social participation/quality of life (COAST) at 6 months. The results for CSLT versus UC and CSLT versus AC comparisons are graphically displayed in forests plots together with the mITT results (*Figures 13–15*). In addition, the interaction tests between the intervention and subgroups are presented in *Table 10*.

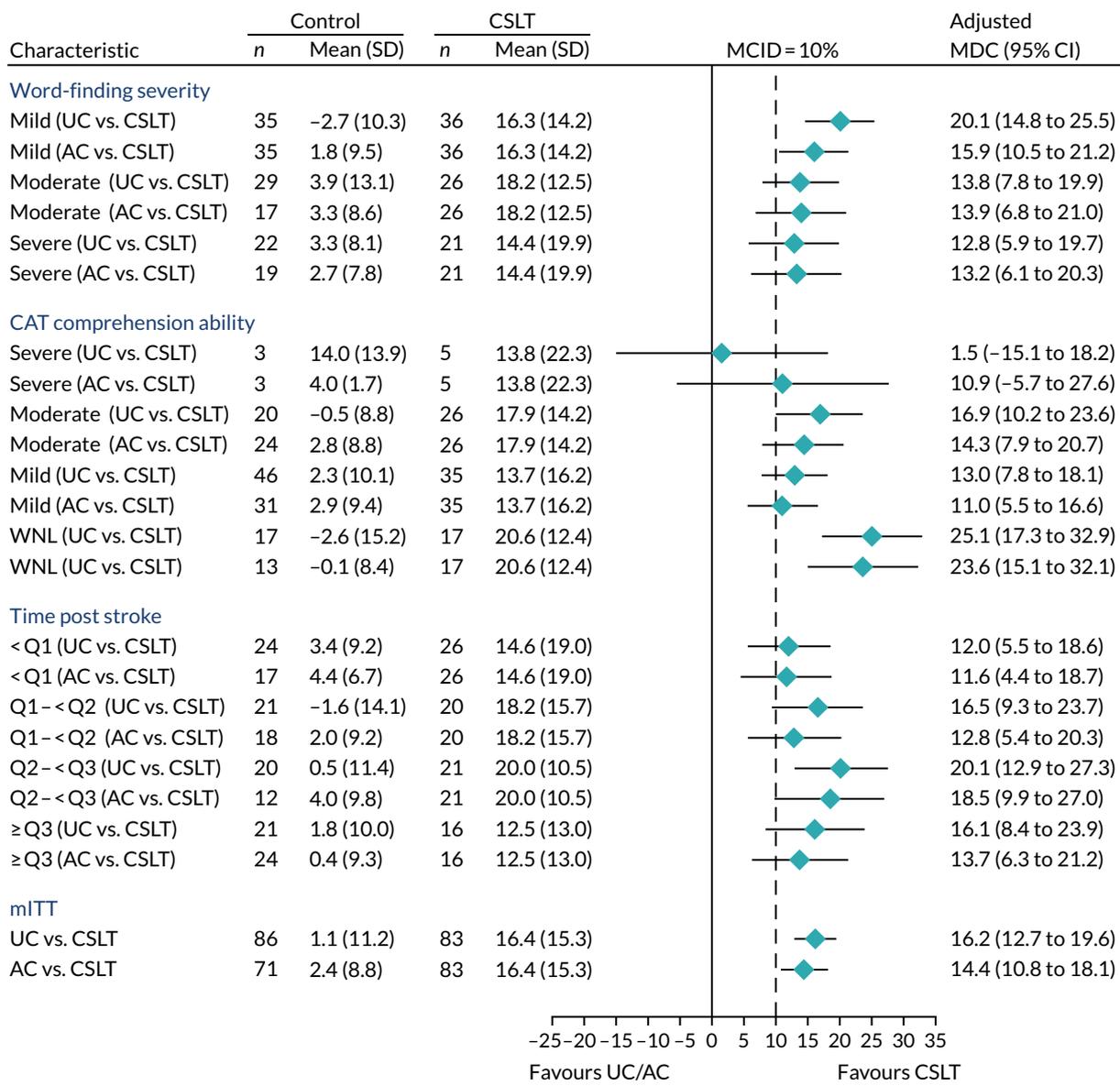


FIGURE 13 Subgroup influence on word-finding results at 6 months. MCID, minimum clinically important difference; Q1, 25th percentile, ≈1 year; Q2, 50th percentile, ≈2 years; Q3, 75th percentile, ≈4 years; WNL, within normal limits.

CLINICAL EFFECTIVENESS RESULTS

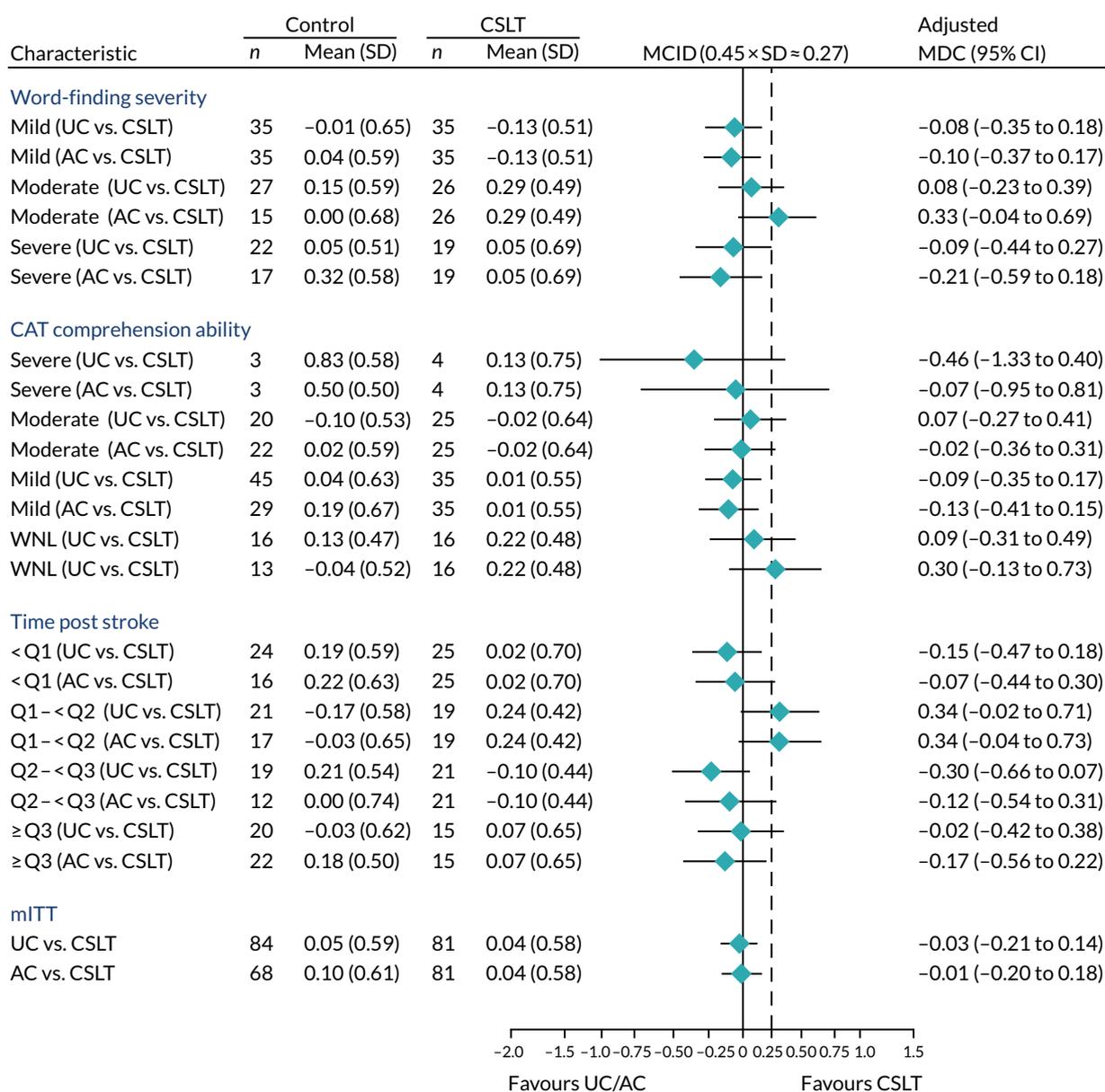


FIGURE 14 Subgroup influence on functional communication results at 6 months. MCID, minimum clinically important difference; Q1, 25th percentile, ≈1 year; Q2, 50th percentile, ≈2 years; Q3, 75th percentile, ≈4 years; WNL, within normal limits.

Due consideration should be given to the clinical or biological plausibility of subgroup results when interpreting these findings. In addition, extreme caution should be taken when interpreting the intervention effect in the severe category of the CAT Comprehension of Spoken Sentences owing to very small sample sizes.

For word-finding, the results appear to be consistent and similar to the mITT results. However, the intervention effect seems to be more pronounced in patients who had mild word-finding difficulties and for those whose comprehension was within normal limits on the CAT Comprehension of Spoken Sentences. The effect of the intervention on TOMs and COAST appears to be broadly consistent across subgroups. However, for unclear clinical reasons, probably due to change, TOMs seem to have improved among the 25% of participants whose time post stroke was between the first and second quartiles of the distribution.

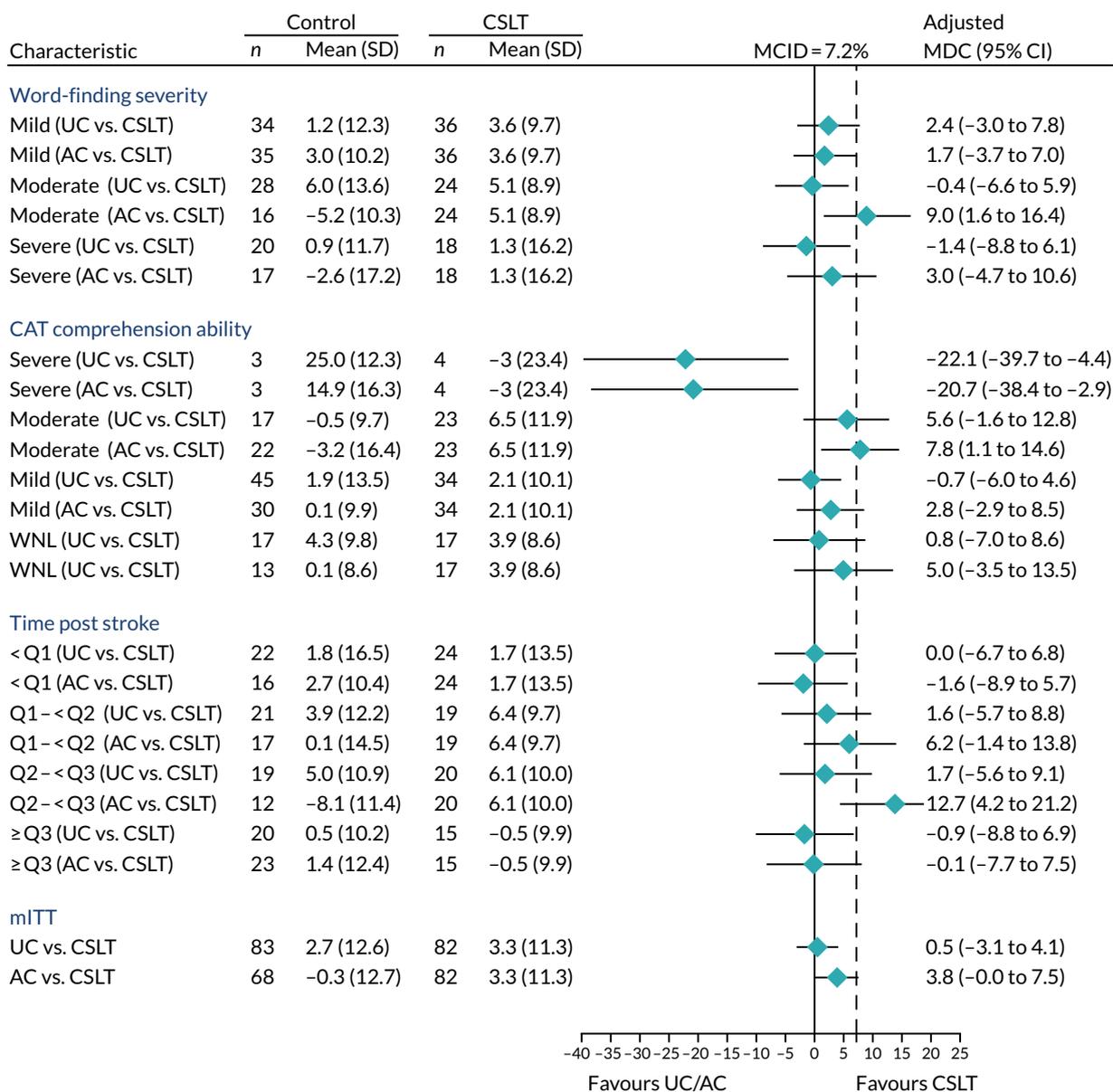


FIGURE 15 Subgroup influence on COAST results at 6 months. MCID, minimum clinically important difference; Q1, 25th percentile, \approx 1 year; Q2, 50th percentile, \approx 2 years; Q3, 75th percentile, \approx 4 years; WNL, within normal limits.

Long-term effects of the intervention on the co-primary and key secondary outcomes

The fourth research objective was to identify whether or not any effects of the interventions were evident 12 months after therapy had begun.

Figures 7–9 show the unadjusted average profile response of participants over time for word-finding, conversation (TOMs) and participant perception of social participation/quality of life (COAST) stratified by the intervention. The effects of the intervention (CSLT vs. UC and CSLT vs. AC) for word-finding, TOMs and COAST at 9 and 12 months are displayed in Figures 16–18, together with the primary results at 6 months (mITT) for comparability.

In summary, the long-term effects of CSLT on word-finding of personal importance at 9 and 12 months were consistent across analysis sets but marginally lower than those observed at 6 months. However, all of the 95% CIs include the 10% improvement in word-finding of clinical relevance and excludes the zero effect of no difference in the intervention effect. Therefore, the sustained long-term intervention effect on word-finding is potentially of clinical importance. For TOMs and COAST, no changes of clinical

TABLE 10 Subgroup analysis results: co-primary and key secondary outcomes at 6 months

Co-primary and key secondary outcomes at 6 months	Subgroup	UC		AC		CSLT		CSLT vs. UC ^a		CSLT vs. AC ^b		
		n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Adjusted MDC (95% CI)	Interaction p-value	Adjusted MDC (95% CI)	Interaction p-value	
Word-finding severity												
Change in word-finding (%) ^c	Mild	35	-2.7 (10.3)	35	1.8 (9.5)	36	16.3 (14.2)	20.1 (14.8 to 25.5)	0.255	15.9 (10.5 to 21.2)	0.828	
	Moderate	29	3.9 (13.1)	17	3.3 (8.6)	26	18.2 (12.5)	13.8 (7.8 to 19.9)		13.9 (6.8 to 21.0)		
	Severe	22	3.3 (8.1)	19	2.7 (7.8)	21	14.4 (19.9)	12.8 (5.9 to 19.7)		13.2 (6.1 to 20.3)		
Change in functional communication ^d	Mild	35	-0.01 (0.65)	35	0.04 (0.59)	35	-0.13 (0.51)	-0.08 (-0.35 to 0.18)	0.88	-0.10 (-0.37 to 0.17)	0.145	
	Moderate	27	0.15 (0.59)	15	0.00 (0.68)	26	0.29 (0.49)	0.08 (-0.23 to 0.39)		0.33 (-0.04 to 0.69)		
	Severe	22	0.05 (0.51)	17	0.32 (0.58)	19	0.05 (0.69)	-0.09 (-0.44 to 0.27)		-0.21 (-0.59 to 0.18)		
Change in COAST (%) ^e	Mild	34	1.2 (12.3)	35	3.0 (10.2)	36	3.6 (9.7)	2.4 (-3.0 to 7.8)	0.809	1.7 (-3.7 to 7.0)	0.322	
	Moderate	28	6.0 (13.6)	16	-5.2 (10.3)	24	5.1 (8.9)	-0.4 (-6.6 to 5.9)		9.0 (1.6 to 16.4)		
	Severe	20	0.9 (11.7)	17	-2.6 (17.2)	18	1.3 (16.2)	-1.4 (-8.8 to 6.1)		3.0 (-4.7 to 10.6)		
CAT comprehension ability												
Change in word-finding (%) ^c	Severe	3	14.0 (3.9)	3	4.0 (1.7)	5	13.8 (22.3)	1.5 (-15.1 to 18.2)	0.034	10.9 (-5.7 to 27.6)	0.271	
	Moderate	20	-0.5 (8.8)	24	2.8 (8.8)	26	17.9 (14.2)	16.9 (10.2 to 23.6)		14.3 (7.9 to 20.7)		
	Mild	46	2.3 (10.1)	31	2.9 (9.4)	35	13.7 (16.2)	13.0 (7.8 to 18.1)		11.0 (5.5 to 16.6)		
	WNL	17	-2.6 (15.2)	13	-0.1 (8.4)	17	20.6 (12.4)	25.1 (17.3 to 32.9)		23.6 (15.1 to 32.1)		
Change in functional communication ^d	Severe	3	0.83 (0.58)	3	0.50 (0.50)	4	0.13 (0.75)	-0.46 (-1.33 to 0.40)	0.587	-0.07 (-0.95 to 0.81)	0.541	
	Moderate	20	-0.10 (0.53)	22	0.02 (0.59)	25	-0.02 (0.64)	0.07 (-0.27 to 0.41)		-0.02 (-0.36 to 0.31)		
	Mild	45	0.04 (0.63)	29	0.19 (0.67)	35	0.01 (0.55)	-0.09 (-0.35 to 0.17)		-0.13 (-0.41 to 0.15)		
	WNL	16	0.13 (0.47)	13	-0.04 (0.52)	16	0.22 (0.48)	0.09 (-0.31 to 0.49)		0.30 (-0.13 to 0.73)		
Change in COAST (%) ^e	Severe	3	25.0 (12.3)	3	14.9 (16.3)	4	-3.1 (23.4)	-22.1 (-39.7 to -4.4)	0.039	-20.7 (-38.4 to -2.9)	0.049	
	Moderate	17	-0.5 (9.7)	22	-3.2 (16.4)	23	6.5 (11.9)	5.6 (-1.6 to 12.8)		7.8 (1.1 to 14.6)		
	Mild	45	1.9 (13.5)	30	0.1 (9.9)	34	2.1 (10.1)	-0.7 (-6.0 to 4.6)		2.8 (-2.9 to 8.5)		
	WNL	17	4.3 (9.8)	13	0.1 (8.6)	17	3.9 (8.6)	0.8 (-7.0 to 8.6)		5.0 (-3.5 to 13.5)		

Co-primary and key secondary outcomes at 6 months	Subgroup	UC		AC		CSLT		CSLT vs. UC ^a		CSLT vs. AC ^b	
		n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Adjusted MDC (95% CI)	Interaction p-value	Adjusted MDC (95% CI)	Interaction p-value
Time post stroke											
Change in word-finding (%) ^c	< Q1	24	3.4 (9.2)	17	4.4 (6.7)	26	14.6 (19.0)	12.0 (5.5 to 18.6)	0.572	11.6 (4.4 to 18.7)	0.647
	Q1-< Q2	21	-1.6 (14.1)	18	2.0 (9.2)	20	18.2 (15.7)	16.5 (9.3 to 23.7)		12.8 (5.4 to 20.3)	
	Q2-< Q3	20	0.5 (11.4)	12	4.0 (9.8)	21	20.0 (10.5)	20.1 (12.9 to 27.3)		18.5 (9.9 to 27.0)	
	≥ Q3	21	1.8 (10.0)	24	0.4 (9.3)	16	12.5 (13.0)	16.1 (8.4 to 23.9)		13.7 (6.3 to 21.2)	
Change in functional communication ^d	< Q1	24	0.19 (0.59)	16	0.22 (0.63)	25	0.02 (0.70)	-0.15 (-0.47 to 0.18)	0.052	-0.07 (-0.44 to 0.30)	0.145
	Q1-< Q2	21	-0.17 (0.58)	17	-0.03 (0.65)	19	0.24 (0.42)	0.34 (-0.02 to 0.71)		0.34 (-0.04 to 0.73)	
	Q2-< Q3	19	0.21 (0.54)	12	0.00 (0.74)	21	-0.10 (0.44)	-0.30 (-0.66 to 0.07)		-0.12 (-0.54 to 0.31)	
	≥ Q3	20	-0.03 (0.62)	22	0.18 (0.50)	15	0.07 (0.65)	-0.02 (-0.42 to 0.38)		-0.17 (-0.56 to 0.22)	
Change in COAST (%) ^e	< Q1	22	1.8 (16.5)	16	2.7 (10.4)	24	1.7 (13.5)	0.0 (-6.7 to 6.8)	0.814	-1.6 (-8.9 to 5.7)	0.047
	Q1-< Q2	21	3.9 (12.2)	17	0.1 (14.5)	19	6.4 (9.7)	1.6 (-5.7 to 8.8)		6.2 (-1.4 to 13.8)	
	Q2-< Q3	19	5.0 (10.9)	12	-8.1 (11.4)	20	6.1 (10.0)	1.7 (-5.6 to 9.1)		12.7 (4.2 to 21.2)	
	≥ Q3	20	0.5 (10.2)	23	1.4 (12.4)	15	-0.5 (9.9)	-0.9 (-8.8 to 6.9)		-0.1 (-7.7 to 7.5)	

Q1, 25th percentile, ≈1 year; Q2, 50th percentile, ≈2 years; Q3, 75th percentile, ≈4 years; WNL, within normal limits.

a UC as the reference group.

b AC as the reference group.

c Higher scores indicate improved vocabulary of personal importance.

d Higher scores indicate improved functional communication ability in conversation.

e Higher percentage score indicates improved patient perception of communication effectiveness and its impact on their quality of life.

Note

Results are based on a multiple linear regression model adjusted for baseline measures and fixed stratification factors (centre and severity of word-finding) with interaction between intervention group and subgroup.

CLINICAL EFFECTIVENESS RESULTS

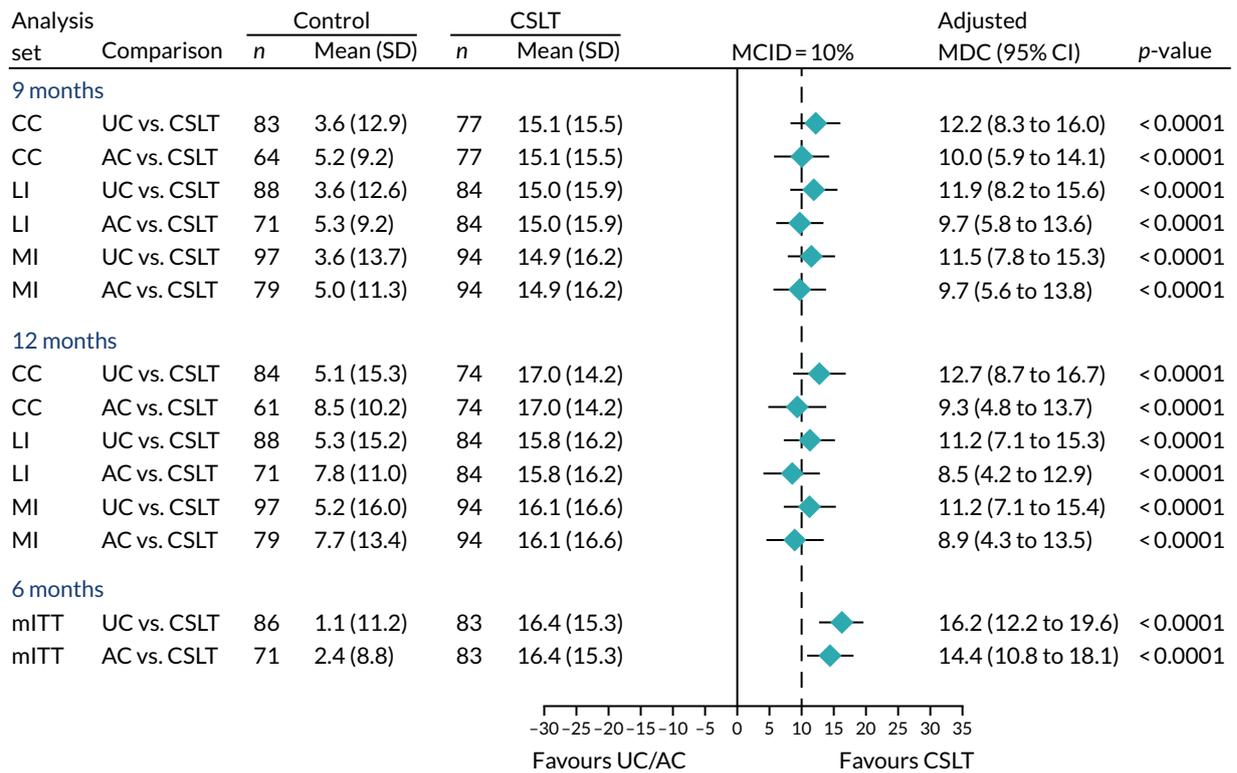


FIGURE 16 Long-term intervention effect on word-finding of personal importance. CC, complete case; LI, linear interpolation; MCID, minimum clinically important difference.

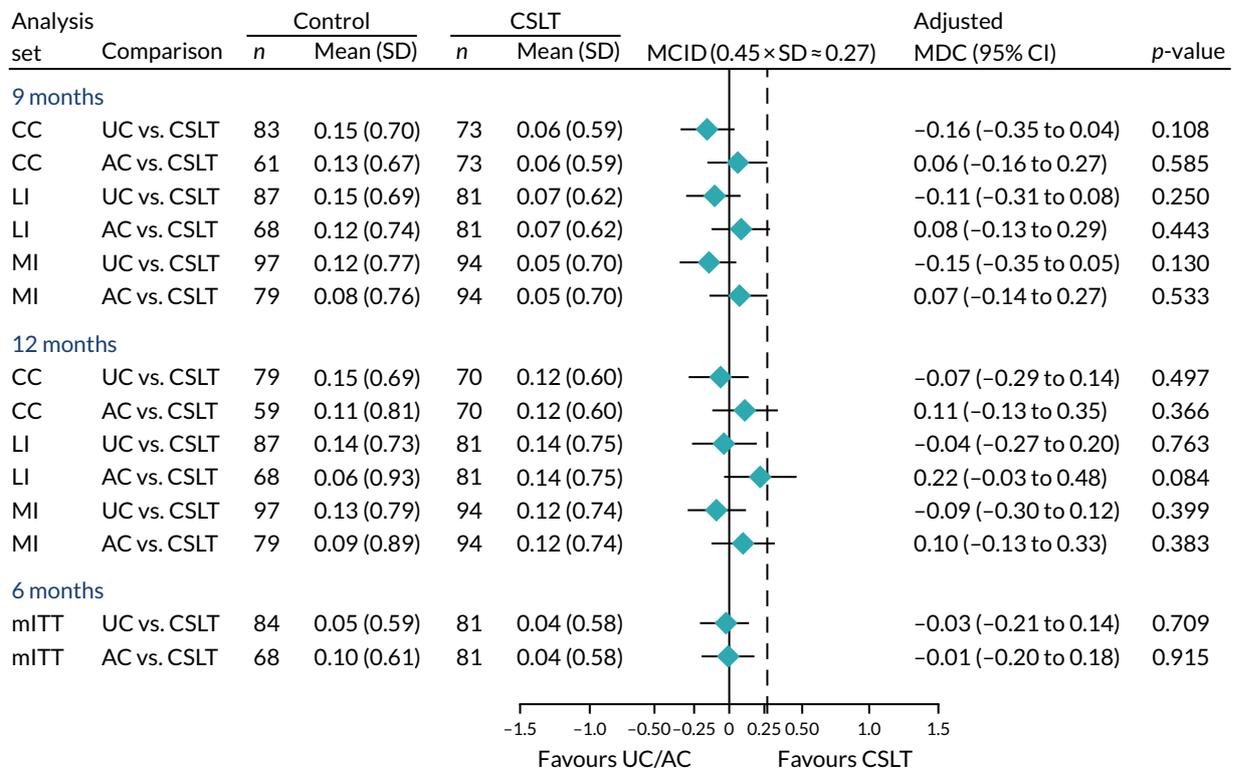


FIGURE 17 Long-term intervention effect on functional communication. CC, complete case; LI, linear interpolation; MCID, minimum clinically important difference.

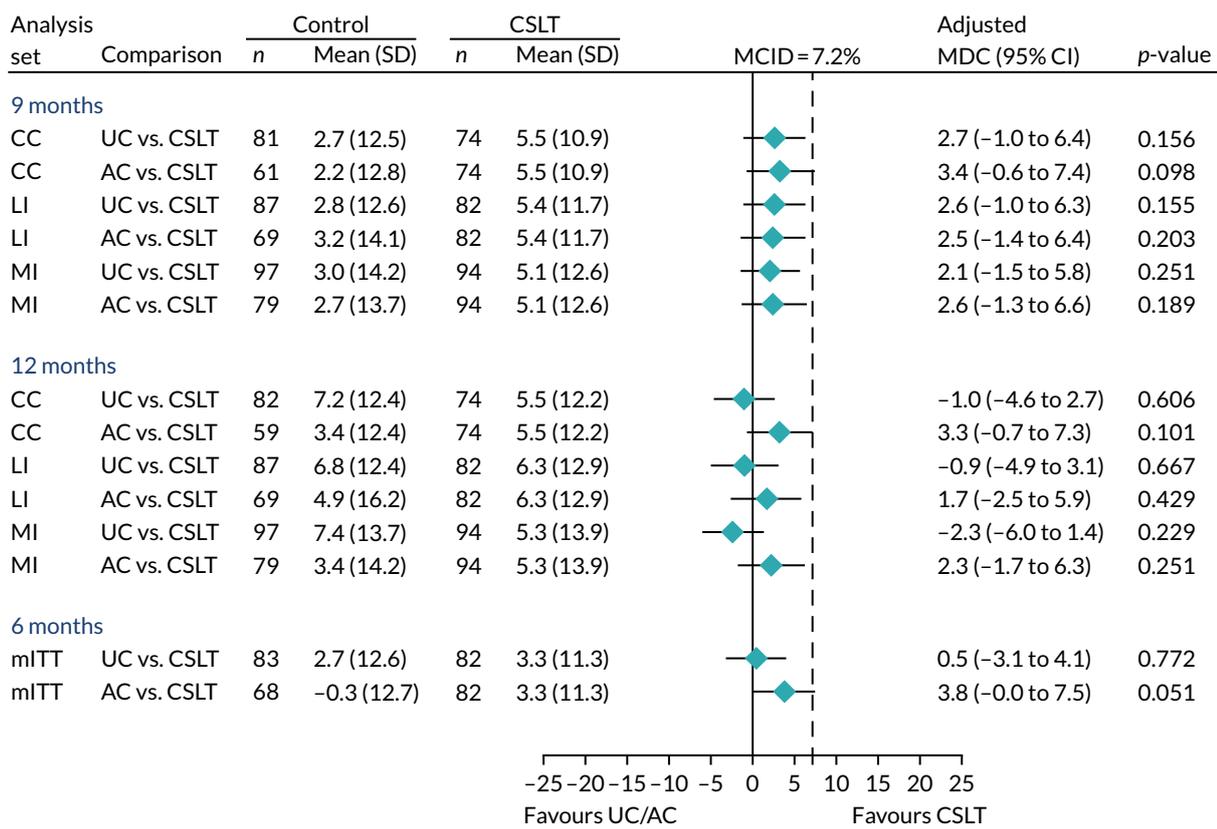


FIGURE 18 Long-term intervention effect on the COAST. CC, complete case; LI, linear interpolation; MCID, minimum clinically important difference.

relevance were observed and the effects are consistent with the primary results at 6 months and across the considered analysis sets.

Effect of the intervention on generalisation to untreated words

In this section, we present the results exploring whether or not there is a generalisation of word-finding from words of personal importance (treated words) to untreated words. This was assessed using the CAT Naming Objects with a possible total score from the picture-naming tasks of 24 untreated words ranging from 0 to 48.

Table 11 summarises the effect of the intervention on word-finding of untreated words at 6, 9 and 12 months from randomisation. At 6 months, the mean change in word-finding of untreated words (scores) was 3.9 (SD 7.9), 0.7 (SD 8.5) and 3.3 (SD 7.0) in the UC, AC and CSLT groups, respectively. This indicates an adjusted MDC of -0.3 (95% CI -2.7 to 2.1) in favour of UC compared with CSLT. On average, the adjusted word-finding of untreated words in the AC group was lower than those of the UC and CSLT groups across assessments.

In summary, there is insufficient evidence to support the positive effect of the intervention on improving the generalisation of word-finding to untreated words in either the short term or long term.

The literature often reports the proportion of participants who do or do not generalise word-finding to untreated words.⁵¹ Therefore, we undertook a post hoc analysis to explore the intervention effect on the proportion of participants who generalised the word-finding of untreated words as defined by a clinical improvement of at least 5% and 10% from baseline. In summary, these results do not support that CSLT increases the proportion of participants achieving a clinical improvement of at least 5% and 10% in the generalisation of word-finding to untreated words compared with UC or AC. Detailed

TABLE 11 Intervention effect on generalisation to untreated words at 6, 9 and 12 months

Change in word-finding of untreated words	UC		AC		CSLT		CSLT vs. UC ^a		CSLT vs. AC ^b		AC vs. UC ^a	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Adjusted MDC (95% CI)	p-value	Adjusted MDC (95% CI)	p-value	Adjusted MDC (95% CI)	p-value
CC												
6 months	86	3.9 (7.9)	69	0.7 (8.5)	82	3.3 (7.0)	-0.3 (-2.7 to 2.1)	0.810	2.6 (0.1 to 5.1)	0.045	-2.9 (-5.3 to -0.4)	0.025
9 months	83	4.8 (8.0)	63	2.5 (6.9)	76	4.0 (7.9)	-0.7 (-3.1 to 1.6)	0.534	1.4 (-1.2 to 3.9)	0.299	-2.1 (-4.6 to 0.4)	0.100
12 months	83	4.5 (8.5)	60	2.8 (7.1)	74	4.8 (7.3)	0.6 (-1.8 to 3.0)	0.634	1.8 (-0.8 to 4.5)	0.177	-1.3 (-3.9 to 1.3)	0.342
MI												
6 months	97	3.9 (8.9)	79	0.8 (8.6)	94	3.4 (7.9)	-0.4 (-2.7 to 2.0)	0.754	2.4 (-0.1 to 4.9)	0.058	-2.8 (-5.3 to -0.3)	0.030
9 months	97	5.0 (8.7)	79	2.4 (7.8)	94	3.9 (8.8)	-1.0 (-3.3 to 1.3)	0.396	1.4 (-1.2 to 3.9)	0.288	-2.4 (-4.9 to 0.1)	0.064
12 months	97	4.8 (9.2)	79	2.3 (8.3)	94	4.7 (8.4)	-0.1 (-2.5 to 2.4)	0.967	2.1 (-0.5 to 4.6)	0.117	-2.1 (-4.7 to 0.5)	0.108
Linear interpolation												
6 months	88	3.8 (7.9)	70	0.7 (8.4)	83	3.3 (7.0)	-0.2 (-2.5 to 2.2)	0.893	2.5 (0.0 to 5.0)	0.047	-2.7 (-5.1 to -0.2)	0.032
9 months	88	4.6 (8.2)	70	2.0 (8.3)	83	4.0 (7.8)	-0.3 (-2.7 to 2.1)	0.808	2.0 (-0.6 to 4.6)	0.131	-2.3 (-4.8 to 0.3)	0.078
12 months	88	4.3 (9.1)	70	2.0 (9.6)	83	4.8 (7.9)	0.8 (-1.9 to 3.4)	0.571	2.7 (-0.2 to 5.5)	0.064	-1.9 (-4.7 to 0.9)	0.179

CC, complete case.

a UC as the reference group.

b AC as the reference group.

Notes

Total score from the picture-naming tasks ranges from 0 to 48. Results are based on a multiple linear regression model adjusted for baseline measures and fixed stratification factors (centre and severity of word-finding).

Positive higher scores indicate improved word-finding of untreated words.

The comparisons of interest are CSLT vs. UC and CSLT vs. AC. AC vs. UC is presented for completeness but shaded to indicate that it is not of clinical importance.

results are accessible online (see section 9.14.1 of the statistical report, which is available online as a supplementary appendix to Palmer *et al.*⁵⁰).

Effect of the intervention on generalisation of treated words used in conversation

Here, we present the results exploring the intervention effect on the use of learned vocabulary (from the word-finding treatment of personally chosen words in conversation). This is based on a possible total number of unique words retrieved in video conversations ranging from 0 to 100.

The average response profiles in the numbers of treated words used in conversation over time are shown in Figure 19. The UC and AC groups experienced an average decrease of about two personally chosen words at 6 months, whereas the CSLT group remained almost the same throughout the trial. This translated to an adjusted MDC in treated words used in conversation at 6 months of only 2.0 (95% CI 0.6 to 3.4) in favour of CSLT compared with UC and 2.9 (95% CI 1.4 to 4.4) in favour of CLST compared with AC. The effect diminished at 9 and 12 months as UC and AC participants improved slightly on average. Table 12 details results that are very consistent across analysis sets considered at 6, 9 and 12 months.

In line with the analysis of generalisation of treated words to untreated words, we performed post hoc analysis exploring the proportion of participants meeting clinical improvement in the generalisation of treated words used in conversation (words retrieved during videoed conversations) of at least five and at least 10 words. As shown in Table 13, 23 (28.4%) participants in the CSLT group, 6 (8.8%) participants in the AC group and 8 (9.5%) participants in the UC group recorded a clinical improvement of at least five words at 6 months from baseline. That is, only about 1 in 10 participants in the UC or AC groups showed a clinical improvement of at least five words, compared with approximately 3 in 10 in the CSLT group. This translates to a 18.9% or 19.6% (about 2 in 10) increase in the proportion of participants showing clinical improvement in the generalisation of treated words used in conversation in the CSLT group compared with the AC or UC groups, respectively, at 6 months. Only a handful of participants improved by > 10 words across interventions in the short and long term at 6, 9 and 12 months.

Carer-rated communication effectiveness and impact on carers' quality of life

One of the additional research objectives is to investigate the effect of the intervention on the carer-rated communication effectiveness (using the first 15 questions from the CaCOAST) and impact on the carers' quality of life (using the last five questions of the CaCOAST). We refer to these as 'CaCOAST 15' and 'CaCOAST 5', respectively. It should be noted that this exploratory analysis includes only available carers who agreed to take part. In addition, the unit of randomisation was the participant and not the carer.

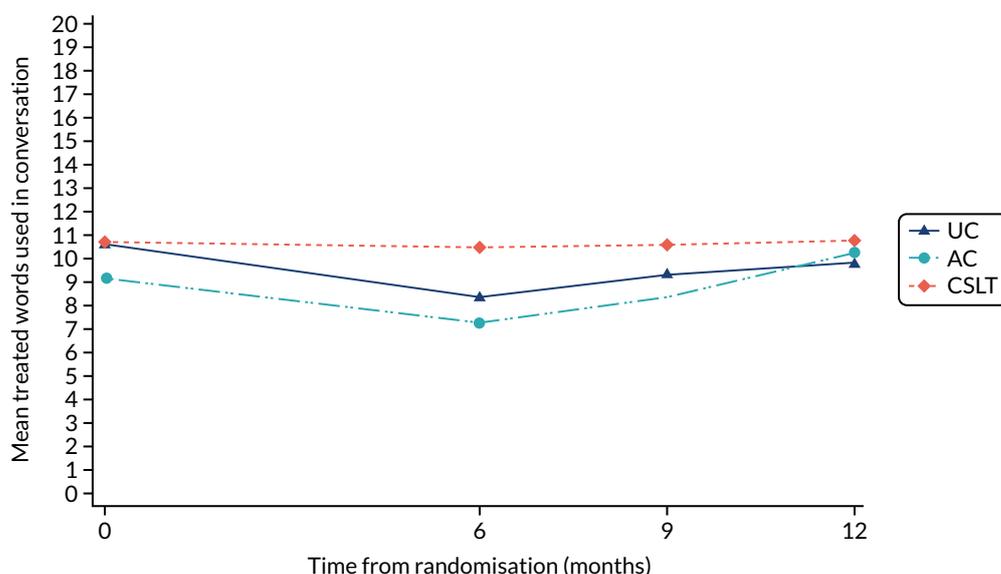


FIGURE 19 Mean response profile in word-finding of treated words over time.

TABLE 12 Intervention effect on generalisation of treated words used in conversation

Change in treated words used in conversation	UC		AC		CSLT		CSLT vs. UC ^a		CSLT vs. AC ^b		AC vs. UC ^a	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Adjusted MDC (95% CI)	p-value	Adjusted MDC (95% CI)	p-value	Adjusted MDC (95% CI)	p-value
CC												
6 months	84	-1.9 (5.3)	68	-2.0 (5.2)	81	0.1 (6.6)	2.0 (0.6 to 3.4)	0.006	2.9 (1.4 to 4.4)	<0.001	-0.9 (-2.4 to 0.6)	0.241
9 months	83	-1.0 (5.6)	61	-1.4 (5.7)	73	0.4 (7.0)	1.4 (-0.3 to 3.0)	0.099	2.2 (0.4 to 4.0)	0.017	-0.8 (-2.6 to 0.9)	0.357
12 months	79	-0.4 (5.5)	59	0.1 (6.6)	70	0.5 (5.6)	0.8 (-0.8 to 2.4)	0.351	0.8 (-1.0 to 2.5)	0.401	0.0 (-1.7 to 1.7)	0.996
MI												
6 months	97	-2.0 (6.0)	79	-1.7 (5.8)	94	-0.3 (7.0)	2.0 (0.5 to 3.5)	0.009	2.8 (1.2 to 4.3)	<0.001	-0.8 (-2.3 to 0.7)	0.317
9 months	97	-1.3 (6.3)	79	-1.2 (6.5)	94	-0.1 (7.4)	1.3 (-0.3 to 2.9)	0.103	2.2 (0.4 to 4.1)	0.017	-0.9 (-2.7 to 0.9)	0.330
12 months	97	-0.4 (6.2)	79	0.5 (6.8)	94	0.3 (6.6)	0.7 (-0.8 to 2.3)	0.354	0.8 (-0.9 to 2.5)	0.347	-0.1 (-1.8 to 1.6)	0.908
Linear interpolation												
6 months	87	-1.9 (5.2)	68	-1.8 (5.3)	81	0.3 (6.8)	2.2 (0.8 to 3.6)	0.003	2.9 (1.4 to 4.4)	<0.001	-0.8 (-2.3 to 0.7)	0.308
9 months	87	-1.1 (5.7)	68	-1.1 (5.7)	81	0.5 (7.4)	1.6 (0.0 to 3.1)	0.047	2.3 (0.7 to 4.0)	0.006	-0.8 (-2.4 to 0.9)	0.361
12 months	87	-0.6 (5.8)	68	0.3 (6.4)	81	0.4 (7.0)	0.8 (-0.8 to 2.5)	0.328	0.9 (-0.9 to 2.7)	0.311	-0.1 (-1.8 to 1.7)	0.915

CC, complete case.

a UC as the reference group.

b AC as the reference group.

Notes

Total numbers of unique words retrieved from the video conversations range from 0 to 100.

Results are based on a multiple linear regression model adjusted for baseline measures and fixed stratification factors (centre and severity of word-finding).

High positive scores indicate improved word-finding of untreated words.

TABLE 13 Clinical improvement in generalisation of treated words used in conversation

Change in treated words used in conversation	Trial group, n (%)			Difference in proportions (95% CI)		
	UC	AC	CSLT	CSLT vs. UC ^a	CSLT vs. AC ^b	AC vs. UC ^a
6 months	N = 84	N = 68	N = 81			
≥ 5 words	8 (9.5)	6 (8.8)	23 (28.4)	18.9 (7.2 to 30.5)	19.6 (7.7 to 31.5)	-0.7 (-9.9 to 8.5)
≥ 10 words	1 (1.2)	0 (0.0)	5 (6.2)	5.0 (-0.7 to 10.7)	6.2 (0.9 to 11.4)	-1.2 (-3.5 to 1.1)
9 months	N = 83	N = 61	N = 73			
≥ 5 words	11 (13.3)	8 (13.1)	15 (20.5)	7.3 (-4.5 to 19.1)	7.4 (-5.1 to 20.0)	-0.1 (-11.3 to 11.0)
≥ 10 words	3 (3.6)	1 (1.6)	4 (5.5)	1.9 (-4.7 to 8.5)	3.8 (-2.3 to 10.0)	-2.0 (-7.1 to 3.2)
12 months	N = 79	N = 59	N = 70			
≥ 5 words	12 (15.2)	12 (20.3)	18 (25.7)	10.5 (-2.4 to 23.5)	5.4 (-9.1 to 19.9)	5.1 (-7.8 to 18.1)
≥ 10 words	3 (3.8)	4 (6.8)	2 (2.9)	-0.9 (-6.7 to 4.8)	-3.9 (-11.4 to 3.6)	3.0 (-4.7 to 10.7)

a UC as the reference group.
b AC as the reference group.

Notes
Complete cases only.
CIs estimated using the normal approximation to the binomial distribution.

The mean change in carer-rated communication effectiveness at 6 months was 6.8% in the CSLT group, compared with 1.0% in the UC group, translating to an adjusted MDC of 4.6% (95% CI 0.3% to 9.0%) in favour of CSLT (Table 14). This small improvement in carer-rated communication effectiveness in CSLT was similar when compared with AC: 5.1% (95% CI 0.5% to 9.7%). However, the long-term effects of the intervention on average change in the carer-rated communication effectiveness were very small: 0.6% (95% CI -4.4% to 5.7%) and 2.7% (95% CI -1.9% to 7.4%) in favour of CSLT compared with UC at 9 and 12 months, respectively.

For the carer-rated impact on their quality of life at 6 months, UC decreased by 1% and CSLT improved by 5.3%, translating to a positive adjusted MDC of 5.3% (95% CI -1.1% to 11.7%) in favour of CSLT. However, although this seems to be of potential clinical relevance, we cannot rule out the lack of benefit. The improvement in carers' quality of life in CSLT compared with AC was close to zero: 0.3% (95% CI -6.4% to 6.9%). In other words, about 5% improvement in carers' quality of life could be attributed the attention given rather than the computer therapy alone. The average effect at 9 months was slightly lower than at 6 months (4.0%, 95% CI -3.3% to 11.2%) in favour of CSLT compared with UC. Although CIs around observed effects include the null effect of zero (no difference between groups), clinical judgements need to be made on whether or not the observed treatment effects are of clinical importance or relevance. In addition, interpretation should be made in consideration of the observed effect between CSLT and AC.

Safety and negative effects

The incidences of negative effects of computer therapy

Table 15 summarises negative effects of computer therapy among the participants who used a computer for the intervention. Of the 85 participants who used computer therapy, 23 (27.1%) felt that the computer practice made them overtired and anxious or worried, translating to an average IR of one episode per person-year. The incidence of computer therapy causing headaches and affecting eyes was low. The results based on treatment-as-received (see Table 15) and treatment-as-randomised (Table 16) principles are very similar.

TABLE 14 Intervention effect on the carer-rated communication effectiveness and carers' quality of life

CaCOAST domain	UC		AC		CSLT		CSLT vs. UC ^a		CSLT vs. AC ^b		AC vs. UC ^a	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Adjusted MDC (95% CI)	p-value	Adjusted MDC (95% CI)	p-value	Adjusted MDC (95% CI)	p-value
Change in CaCOAST 15 (%)^c												
6 months	48	1.0 (11.5)	38	2.6 (11.3)	51	6.8 (11.5)	4.6 (0.3 to 9.0)	0.038	5.1 (0.5 to 9.7)	0.030	-0.5 (-5.2 to 4.2)	0.846
9 months	43	2.5 (11.7)	37	4.4 (11.3)	47	5.7 (13.2)	0.6 (-4.4 to 5.7)	0.802	0.5 (-4.7 to 5.6)	0.855	0.2 (-5.1 to 5.5)	0.953
12 months	45	3.1 (11.5)	34	4.2 (11.3)	44	6.6 (11.6)	2.7 (-1.9 to 7.4)	0.244	1.9 (-3.1 to 6.9)	0.444	0.8 (-4.2 to 5.8)	0.752
Change in CaCOAST 5 (%)^d												
6 months	48	-1.0 (15.8)	38	7.2 (14.6)	51	5.8 (17.3)	5.3 (-1.1 to 11.7)	0.105	0.3 (-6.4 to 6.9)	0.940	5.0 (-2.0 to 12.0)	0.156
9 months	43	-0.6 (18.4)	37	8.2 (14.6)	47	8.5 (18.7)	4.0 (-3.3 to 11.2)	0.279	0.2 (-7.0 to 7.5)	0.949	3.7 (-4.0 to 11.4)	0.337
12 months	45	4.0 (20.1)	34	7.2 (13.6)	44	8.5 (19.5)	0.6 (-6.3 to 7.4)	0.871	3.4 (-4.0 to 10.8)	0.363	-2.8 (-10.4 to 4.8)	0.460

CC, complete case.

a UC as the reference group.

b AC as the reference group.

c High positive scores indicate improved carer-rated communication effectiveness.

d High positive scores indicate improved impact on carers' quality of life.

Note

Results based on a multiple linear regression model adjusted for baseline measures and fixed stratification factors (centre and severity of word-finding); only for CC set.

TABLE 15 Negative effects of computer therapy (only those who used the computer)

Has the computer practice:	CSLT (N = 85)		
	n (%)	Total events/person-years	IR/person-year (95% CI)
Made you feel overtired?	23 (27.1)	50/42.1	1.18 (0.74 to 1.90)
Affected your eyes?	11 (12.9)	17/42.1	0.40 (0.21 to 0.78)
Given you headaches?	5 (5.9)	6/42.1	0.14 (0.06 to 0.36)
Made you feel anxious/worried?	23 (27.1)	42/42.1	0.99 (0.63 to 1.56)

Results from a negative binomial regression model.

TABLE 16 Negative effects of the computer therapy (treatment as randomised)

Has the computer practice:	CSLT (N = 97)		
	n (%)	Total events/person-years	IR/person-year (95% CI)
Made you feel over tired?	26 (26.8)	53/47.1	1.14 (0.73 to 1.78)
Affected your eyes?	11 (11.3)	17/47.1	0.36 (0.18 to 0.70)
Given you headaches?	5 (5.2)	6/47.1	0.13 (0.05 to 0.32)
Made you feel anxious/worried?	26 (26.8)	45/47.1	0.97 (0.63 to 1.49)

Results from a negative binomial regression model.

The incidences of adverse events

Of the 97 participants randomised to receive CSLT, computer therapy use data were not recorded for 12 (12.4%), three of whom died before the 6-month assessment. Only one participant who was allocated to AC died before the 6 months and was never sent a puzzle book. Therefore, in accordance with the treatment-as-received principle, these 13 participants were technically treated as having received UC alone. The other four participants who died were allocated to the UC group and as having not received any other treatments. Thus, the denominators for the treatment-as-received analysis are 101 in the UC group, 79 in the AC group and 85 in the CSLT group.

Table 17 summarises the incidences of AEs using the treatment-as-received principle. The proportion of participants who experienced any AE was 61 (71.8%) in the CSLT group, 50 (63.3%) in the AC group and 70 (61.4%) in the UC group. On average, the number of AEs per participant per person-year of follow-up was 2.18, 1.79 and 1.87 in the CSLT, AC and UC groups, respectively. This indicates a slight increase in all AEs in the CSLT group, with an IRR of 1.16 (95% CI 0.83 to 1.62) and 1.22 (95% CI 0.85 to 1.77) compared with the UC and AC groups, respectively. However, there is insufficient evidence to suggest an increased risk of AEs in the CSLT group compared with the AC or UC groups.

Although fits and seizures were not common, the risk of these AEs was more than three times higher in the CSLT group than in either the UC group or the AC group. The numbers of AEs by category are also presented in Table 17. Unfortunately, we did not record if participants had had previous seizures before taking part in the trial; however, the events were all considered unrelated or unlikely to be related to the intervention by the clinicians at the sites.

For sensitivity analysis, Table 18 summarises AEs using the strict intention-to-treat (ITT) principle (treatment as randomised). In summary, the interpretation is consistent with the treatment-as-received results presented in Table 17.

CLINICAL EFFECTIVENESS RESULTS

TABLE 17 Incidence of AEs (treatment as received)

AE classification	UC (N = 114)	AC (N = 79)	CSLT (N = 85)	IRR (95% CI)		
				CSLT vs. UC ^a	CSLT vs. AC ^b	AC vs. UC ^a
Had experienced at least one AE, n (%)	70 (61.4)	50 (63.3)	61 (71.8)			
Repeated AEs						
<i>All AEs</i>						
Total events/person-years	200/105.4	136/74.7	185/84.7			
IR/person-year (95% CI)	1.87 (1.47 to 2.38)	1.79 (1.38 to 2.31)	2.18 (1.72 to 2.77)	1.16 (0.83 to 1.62)	1.22 (0.85 to 1.77)	0.95 (0.67 to 1.35)
<i>Felt more tired than usual</i>						
Total events/person-years	125/105.4	77/74.7	114/84.7			
IR/person-year (95% CI)	1.18 (0.82 to 1.70)	1.01 (0.70 to 1.45)	1.32 (0.95 to 1.84)	1.12 (0.69 to 1.83)	1.32 (0.81 to 2.14)	0.85 (0.51 to 1.42)
<i>Had any fits or seizures</i>						
Total events/person-years	18/105.4	13/74.7	47/84.7			
IR/person-year (95% CI)	0.16 (0.06 to 0.44)	0.17 (0.08 to 0.37)	0.57 (0.29 to 1.12)	3.48 (1.05 to 11.57)	3.41 (1.21 to 9.62)	1.02 (0.29 to 3.63)
<i>Had worsening vision or visual difficulties</i>						
Total events/person-years	47/105.4	34/74.7	71/84.7			
IR/person-year (95% CI)	0.42 (0.22 to 0.80)	0.44 (0.25 to 0.79)	0.83 (0.51 to 1.36)	1.95 (0.87 to 4.37)	1.89 (0.89 to 4.05)	1.03 (0.43 to 2.44)
<i>Had increasing number or increasing severity of headaches</i>						
Total events/person-years	46/105.4	25/74.7	52/84.7			
IR/person-year (95% CI)	0.43 (0.23 to 0.81)	0.31 (0.13 to 0.78)	0.58 (0.34 to 1.01)	1.36 (0.59 to 3.11)	1.84 (0.64 to 5.30)	0.74 (0.24 to 2.21)
<i>Had any accidents (e.g. falls) or injuries</i>						
Total events/person-years	90/105.4	51/74.7	48/84.7			
IR/person-year (95% CI)	0.87 (0.58 to 1.30)	0.66 (0.42 to 1.04)	0.56 (0.35 to 0.89)	0.64 (0.35 to 1.19)	0.85 (0.45 to 1.61)	0.76 (0.42 to 1.39)
<i>Reported any other negative effects or events</i>						
Total events/person-years	64/105.4	29/74.7	44/84.7			
IR/person-year (95% CI)	0.60 (0.40 to 0.92)	0.38 (0.21 to 0.68)	0.55 (0.35 to 0.86)	0.91 (0.49 to 1.68)	1.44 (0.69 to 3.00)	0.63 (0.31 to 1.28)

a UC as the reference group.

b AC as the reference group.

Notes

Results are from a negative binomial regression model.

The column AC vs. UC is exploratory, and therefore shaded, but presented for completeness.

The comparisons of interest are CSLT vs. UC and CSLT vs. AC. AC vs. UC is presented for completeness.

Adapted from Palmer *et al.*⁵⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

TABLE 18 Incidence of AEs (treatment as randomised)

AE classification	UC (N = 101)	AC (N = 80)	CSLT (N = 97)	IRR (95% CI)		
				CSLT vs. UC ^a	CSLT vs. AC ^b	AC vs. UC ^a
Had experienced at least one AE, n (%)	62 (61.4)	50 (62.5)	69 (71.1)			
Repeated AEs						
All AEs						
Total events/person-years	186/97.1	136/74.8	199/92.9			
IR/person-year (95% CI)	1.90 (1.46 to 2.48)	1.78 (1.38 to 2.30)	2.13 (1.71 to 2.65)	1.11 (0.80 to 1.56)	1.19 (0.83 to 1.71)	0.93 (0.65 to 1.34)
Felt more tired than usual						
Total events/person-years	122/97.1	77/74.8	117/92.9			
IR/person-year (95% CI)	1.28 (0.84 to 1.93)	1.00 (0.69 to 1.47)	1.21 (0.85 to 1.74)	0.95 (0.56 to 1.61)	1.21 (0.69 to 2.14)	0.78 (0.45 to 1.38)
Had any fits or seizures						
Total events/person-years	15/97.1	13/74.8	50/92.9			
IR/person-year (95% CI)	0.15 (0.04 to 0.52)	0.17 (0.07 to 0.41)	0.53 (0.24 to 1.19)	3.57 (1.01 to 12.67)	3.22 (0.83 to 12.42)	1.11 (0.27 to 4.54)
Had worsening vision or visual difficulties						
Total events/person-years	46/97.1	34/74.8	72/92.9			
IR/person-year (95% CI)	0.45 (0.21 to 0.97)	0.44 (0.23 to 0.84)	0.76 (0.42 to 1.36)	1.69 (0.69 to 4.13)	1.74 (0.67 to 4.57)	0.97 (0.37 to 2.56)
Had increasing number or increasing severity of headaches						
Total events/person-years	44/97.1	25/74.8	54/92.9			
IR/person-year (95% CI)	0.46 (0.21 to 1.00)	0.31 (0.11 to 0.91)	0.54 (0.27 to 1.09)	1.17 (0.39 to 3.52)	1.71 (0.52 to 5.62)	0.69 (0.21 to 2.26)
Had any accidents (e.g. falls) or injuries						
Total events/person-years	86/97.1	51/74.8	52/92.9			
IR/person-year (95% CI)	0.92 (0.58 to 1.45)	0.66 (0.41 to 1.06)	0.55 (0.35 to 0.87)	0.60 (0.32 to 1.13)	0.83 (0.42 to 1.66)	0.72 (0.37 to 1.40)
Reported any other negative effects or events						
Total events/person-years	59/97.1	29/74.8	49/92.9			
IR/person-year (95% CI)	0.61 (0.38 to 0.99)	0.38 (0.20 to 0.71)	0.55 (0.35 to 0.86)	0.90 (0.46 to 1.77)	1.45 (0.68 to 3.08)	0.62 (0.30 to 1.30)
a UC as the reference group.						
b AC as the reference group.						
Notes						
Results are from a negative binomial regression model.						
The comparisons of interest are CSLT vs. UC and CSLT vs. AC. AC vs. UC is presented for completeness.						

The incidence of serious adverse events

The incidence of SAEs based on the treatment-as-received principle is summarised in *Table 19*. The number of participants who experienced any SAEs was 18 (15.8%) in the UC group, 11 (13.9%) in the AC group and nine (10.6%) in the CSLT group. The total number of repeated SAEs was 23, 12 and 10 in the UC, AC and CSLT groups, respectively, which were experienced over a total follow-up of 105.4, 74.7 and 84.7 person-years, respectively. The incidence of SAEs was low across interventions such that participants would need to be followed up for a longer duration to record a single event per participant on average. For instance, the IR in the CSLT group was 0.11 (95% CI 0.04 to 0.19), meaning that, on average, a participant will need to be followed up for about 10 person-years to record one SAE. Although the risk of experiencing any SAEs was lower in the CSLT group than in either the UC group or the AC group, there is insufficient evidence to suggest differences in risk between groups. All SAEs were not related or unlikely to be related to the trial activity and the majority resulted in inpatient hospitalisation.

TABLE 19 Incidence of SAEs (treatment as received)

SAE classification	UC (N = 114)	AC (N = 79)	CSLT (N = 85)	IRR (95% CI)		
				CSLT vs. UC ^a	CSLT vs. AC ^b	AC vs. UC ^a
Had experienced at least one SAE, n (%)	18 (15.8)	11 (13.9)	9 (10.6)			
Repeated SAEs						
<i>All SAEs</i>						
Total events/person-years	23/105.4	12/74.7	10/84.7			
IR/person-year (95% CI)	0.23 (0.11 to 0.34)	0.16 (0.06 to 0.26)	0.11 (0.04 to 0.19)	0.51 (0.22 to 1.19)	0.72 (0.28 to 1.87)	0.70 (0.31 to 1.59)
SAE resulted in inpatient hospitalisation (n)						
No	4	1	0			
Yes	19	11	10			
SAE was life-threatening (n)						
No	14	8	7			
Yes	9	4	3			
Expected						
No	21	12	9			
Yes	1	0	1			
Not stated	1	0	0			
Relationship to trial activity (n)						
Unlikely	1	2	2			
Unrelated	22	10	8			
Frequency of SAE (n)						
Isolated	16	9	7			
Intermittent	2	1	0			
Continuous	3	0	0			
Unknown	2	2	3			

TABLE 19 Incidence of SAEs (treatment as received) (continued)

SAE classification	UC (N = 114)	AC (N = 79)	CSLT (N = 85)	IRR (95% CI)		
				CSLT vs. UC ^a	CSLT vs. AC ^b	AC vs. UC ^a
Intensity of SAE (n)						
Mild	3	1	1			
Moderate	12	6	5			
Severe	8	4	4			
Missing	0	1	0			
Outcome of SAE (n)						
Recovered	10	3	4			
Improved	3	3	1			
Ongoing	5	5	3			
Death	5	1	2			
Action taken (n)						
None	19	10	9			
Reduce intervention	1	0	0			
Intervention withdrawal	1	1	1			
Other	2	1	0			
<p>a UC as the reference group. b AC as the reference group.</p> <p>Notes Results are from a negative binomial regression model. The comparisons of interest are CSLT vs. UC and CSLT vs. AC. AC vs. UC is presented for completeness.</p>						

Table 20 presents the incidences of SAEs based on the strict ITT principle (treatment as randomised). In summary, although the incidence is now slightly higher in the CSLT group than in the AC or UC groups, the conclusion is similar to that based on results in Table 19 (using treatment as received). That is, there is insufficient evidence to suggest differences in IRs of SAEs across interventions. Appendix 3, Table 29, shows the SAEs by treatment received in categories.

Conclusions

This chapter presented detailed trial results relating to the clinical effectiveness objectives. We explored the impact of missing data and adherence to the intervention on the results as part of sensitivity analysis. In the spirit of good practice and transparency, the statistical analysis methods were pre-planned and documented in the SAP, which is accessible online (www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#). All post hoc analyses have been declared with rationale. The detailed statistical report is also available online as a supplementary appendix to Palmer *et al.*⁵⁰ A detailed discussion of the results with strength, limitations and implications are provided in Chapter 6.

CLINICAL EFFECTIVENESS RESULTS

TABLE 20 Incidence of SAEs (treatment as randomised)

SAE classification	UC (N = 101)	AC (N = 80)	CSLT (N = 97)	IRR (95% CI)		
				CSLT vs. UC ^a	CSLT vs. AC ^b	AC vs. UC ^a
Had experienced at least one SAE, n (%)	13 (12.9)	11 (13.8)	14 (14.4)			
Repeated SAEs						
<i>All SAEs</i>						
Total events/person-years	15/97.1	12/74.8	18/92.9			
IR/person-year (95% CI)	0.16 (0.06 to 0.25)	0.16 (0.05 to 0.26)	0.19 (0.09 to 0.30)	1.24 (0.56 to 2.76)	1.23 (0.52 to 2.88)	1.01 (0.42 to 2.43)
SAE resulted in inpatient hospitalisation (n)						
No	3	1	1			
Yes	12	11	17			
SAE was life-threatening (n)						
No	8	8	13			
Yes	7	4	5			
Expected (n)						
No	13	12	17			
Yes	1	0	1			
Not stated	1	0	0			
Relationship to trial activity (n)						
Unlikely	0	2	3			
Unrelated	15	10	15			
Frequency of SAE (n)						
Isolated	11	9	12			
Intermittent	2	1	0			
Continuous	1	0	2			
Unknown	1	2	4			
Intensity of SAE (n)						
Mild	2	1	2			
Moderate	7	6	10			
Severe	6	4	6			
Missing	0	1	0			
Outcome of SAE (n)						
Recovered	5	3	9			
Improved	3	3	1			
Ongoing	3	5	5			
Death	4	1	3			

TABLE 20 Incidence of SAEs (treatment as randomised) (continued)

SAE classification	UC (N = 101)	AC (N = 80)	CSLT (N = 97)	IRR (95% CI)		
				CSLT vs. UC ^a	CSLT vs. AC ^b	AC vs. UC ^a
Action taken (n)						
None	13	10	15			
Reduce intervention	0	0	1			
Intervention withdrawal	1	1	1			
Other	1	1	1			

a UC as the reference group.

b AC as the reference group.

Notes

Results are from a negative binomial regression model.

The comparisons of interest are CSLT vs. UC and CSLT vs. AC. AC vs. UC is presented for completeness.

Adapted from Palmer *et al.*⁵⁰ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>.

In summary, we demonstrated overwhelming evidence that the CSLT intervention improves the mean ability of people with aphasia to retrieve vocabulary of personal importance in a confrontation naming test (16.2%, 95% CI 12.7% to 19.6%; $p < 0.0001$) compared with UC. These results are strongly supported by a 14.4% (95% CI 10.8% to 18.1%) average improvement in word-finding of personally selected words in favour of CSLT compared with AC. Most importantly, the short-term effect of the intervention was sustained in the long term, at 9 and 12 months after therapy has begun. However, on average, the CSLT intervention for word-finding did not result in improvement in functional communication ability in conversation or changes in perceived social participation in daily activities and quality of life in both the short term and long term.

As for the additional research objectives, on average CSLT did not result in the improved generalisation of treatment to the finding of untreated words. The intervention resulted in small short-term improvement in carers' perception of communication effectiveness, which could be of potential clinical relevance: mean 4.6% (95% CI 0.3% to 9.0%). However, the intervention failed to translate into meaningful impact on the carers' reported quality of life based on the perceptions of the carers who were available and agreed to take part in the trial.

As for safety objectives, the incidences of AEs and SAEs were similar across interventions and were generally low. The most negative effects of the computer therapy recorded by 27% of the CSLT participants were feelings of overtiredness and anxiety.

Chapter 5 Health economics

Health economic analysis: summary of key points

A summary of the key health economic findings is presented below to help orientate the reader to the detailed explanation in the chapter:

- Our best estimates suggest that CSLT does not represent a cost-effective use of health-care resources for the population included in Big CACTUS, given a cost-effectiveness threshold of £30,000 per QALY gained. The base-case model-based cost-effectiveness analysis resulted in an ICER of £42,686 per QALY gained for CSLT compared with UC [incremental cost £732.73, 95% credible interval (CrI) £674.23 to £798.05; incremental QALYs 0.017, 95% CrI -0.05 to 0.10] and an ICER of £40,164 per QALY gained for CSLT compared with AC (incremental cost £694.59, 95% CrI £636.46 to £760.09; incremental QALYs 0.017, 95% CrI -0.05 to 0.10). AC was dominated by UC. These ICERs are higher than the NICE cost-effectiveness threshold of £20,000–30,000 per QALY gained. At a cost-effectiveness threshold of £30,000 per QALY gained, the probability of CSLT representing the most cost-effective option is 0.32, compared with 0.45 for UC and 0.22 for AC.
- We estimate that CSLT may represent a cost-effective use of health-care resources for people with mild or moderate word-finding difficulty. The ICER for CSLT compared with UC was £22,371 per QALY gained in the mild word-finding difficulty subgroup (incremental cost £653.49, 95% CrI £586.44 to £728.36; incremental QALYs 0.029, 95% CrI -0.06 to 0.17) and was £28,898 per QALY gained in the moderate word-finding difficulty subgroup (incremental cost £822.77, 95% CrI £715.54 to £942.22; incremental QALYs 0.028, 95% CrI -0.10 to 0.22). The probability of CSLT, UC and AC being cost-effective at a threshold of £30,000 per QALY gained was 0.34, 0.32 and 0.34, respectively, in the mild word-finding subgroup, and was 0.41, 0.37 and 0.22, respectively, in the moderate word-finding subgroup.
- The results of the economic analyses were highly sensitive to the approach used to estimate utility scores. In our base case, we estimate utility scores using an unofficial accessible version of the EQ-5D-5L. A more standard approach may have been to use utility scores collected by proxy from the carers of trial participants. Using this approach, the ICER for the complete trial population reduced from £42,686 per QALY gained to £28,819 per QALY gained for the comparison of CSLT and UC. In this scenario, at a cost-effectiveness threshold of £30,000 per QALY gained, the probability of CSLT representing the most cost-effective option was 0.42, compared with 0.39 for UC and 0.19 for AC.
- The cost of the CSLT intervention was low, and therefore very small increases in the estimated QALY gain can have a large impact on the ICER.
- CSLT substantially increased the proportion of participants who achieved a 'good response', compared with UC and AC, as defined by an increase of $\geq 10\%$ in treated words of personal relevance named correctly, or an increase of ≥ 0.5 points in the activity dimension of the TOMs rating scale. However, the quality-of-life impact of a good response – as measured by the EQ-5D-5L – was small and highly uncertain, to the extent that we are very uncertain about whether or not a good response results in any improvement in quality of life. In the base case, the change in utility associated with achieving a good response to treatment was not statistically significant at any of the 6-, 9- and 12-month time points. At the 6- and 9-month time points, achieving a good response was associated with a statistically non-significant reduction in utility; at 12 months, achieving a good response was associated with a statistically non-significant increase in utility. These utility estimates are crucial determinants of the results of the economic evaluation, but are highly uncertain, leading inevitably to highly uncertain cost-effectiveness conclusions. In particular, although the CrI around the incremental QALY gain associated with CSLT is reasonably narrow, it crosses zero. Given its low cost, only a small QALY gain is required in order for CSLT to appear cost-effective, but we are uncertain about whether or not the intervention leads to increased QALYs.

- In the subgroup analyses, the change in utility associated with achieving a good response at 12 months was marginally higher in the mild and moderate word-finding difficulty subgroups than in the severe word-finding difficulty subgroup, leading to more favourable cost-effectiveness estimates. However, these utility improvements were also highly uncertain (and achieving a good response was again associated with a statistically non-significant reduction in utility at earlier time points), and therefore a high degree of uncertainty remains in the cost-effectiveness analyses of these subgroups.
- Uncertain findings on the impact of CSLT on quality of life are consistent with findings on the co-primary outcome of functional communication ability in conversation and the key secondary outcome of patient perception of their communication and its impact on their life, for which CSLT did not show an effect. Research into the translation of word-finding improvements to conversation and other functional contexts that may lead to quality-of-life benefits may be valuable.

Background

Economic evaluation plays an important role in health technology assessment around the world, with the aim of ensuring that limited health-care budgets are allocated efficiently. The trial was set in the UK, where NICE produces guidance on clinical practice and the use of health technologies in the NHS. NICE produces a *Guide to the Methods of Technology Appraisal*⁵² and the analyses presented here follow that and other good-practice guides,⁵²⁻⁵⁶ supplemented with additional analyses that are not usually considered by NICE (for instance, we present a supplementary analysis including volunteer costs, representing a broader perspective than that associated with only NHS and PSS costs).

The objective of the health economic analysis is to evaluate evidence on the cost-effectiveness of self-managed CSLT, AC and UC for patients with persistent aphasia post stroke. The population represents that included in the Big CACTUS trial, and subgroup analyses were undertaken on the prespecified trial subgroups (see *Subgroup analysis* for details). This work builds on a previous economic evaluation conducted alongside the CACTUS study, a pilot study that led to the Big CACTUS trial.¹⁸

Overview of health economics methods

Health economics data on resource use, costs and health-related quality-of-life (HRQoL) scores associated with the CSLT, AC and UC treatment groups were collected alongside the trial. The HRQoL data were collected using the EQ-5D-5L, allowing QALYs to be calculated, and therefore a series of cost-utility analyses were conducted, expressing results in terms of incremental cost per QALY gained. For context, NICE usually considers interventions to represent a cost-effective use of NHS resources if the ICER is less than £20,000–30,000 per QALY gained.⁵²

The primary analysis consisted of a model-based analysis with a lifetime time horizon that adopted an NHS and PSS perspective. This is in line with NICE requirements for interventions that may have benefits and/or costs that persist beyond the duration of the clinical trial.⁵² We also present a secondary within-trial analysis to illustrate the benefits/costs incurred within the trial duration, which avoids the extrapolation necessary for the model-based analysis. In addition, owing to the importance of volunteers who helped some participants with their use of the CSLT intervention in the Big CACTUS trial, we present supplementary model-based and within-trial analyses that adopt a broader perspective, allowing costs incurred by these volunteers to be incorporated. This does not represent a complete societal perspective; for example, patient and carer time costs and costs related to participants' own computers were not included. However, given that volunteers were used to help implement the intervention, it is useful to investigate how these costs might influence the economic evaluation.

For all analyses, a fully incremental analysis was undertaken in which incremental costs and QALYs were calculated for each of the three treatment options ranked by ascending cost. This allows all possible cost-effectiveness comparisons to be made (i.e. CSLT vs. UC, CSLT vs. AC, AC vs. UC). However, the comparison of primary interest for the economic evaluation is CSLT vs. UC, representing what are perceived to be the two viable/most likely treatment options to be used in clinical practice.

Methods: model-based analysis (primary analysis)

Model design

A Markov model was developed to estimate the long-term cost-effectiveness of the CSLT intervention for the primary analysis using Microsoft Excel® 2010 (Microsoft Corporation, Redmond, WA, USA). The structure of the model was adapted from a previous model developed alongside the pilot CACTUS study.¹⁸ The previously developed model consisted of three health states:

1. aphasia
2. good response
3. dead.

Owing to the additional time points at which data were collected in the Big CACTUS trial, it was possible to differentiate between responses achieved (or maintained) at the 6-, 9- and 12-month assessments. Therefore, tunnel states (i.e. states in which participants cannot remain for more than one modelled cycle) representing 'good response (6 months)' and 'good response (9 months)' were added to the model. Whether or not a patient achieved a good response was based on the co-primary outcome measures of the Big CACTUS trial: relating to change in word-finding ability and change in functional communication in conversation (see *Chapter 3, Word-finding of personally selected words for treatment and Functional communication in conversation* for descriptions of these measures). In the economic model, participants moved into a 'good response' health state if they achieved the defined minimum change for either one of these measures, where the minimum change was an increase of $\geq 10\%$ in treated words of personal relevance named correctly for the word-finding outcome, or an increase of ≥ 0.5 points since baseline in the activity dimension of the TOMs rating scale for the functional outcome. Because we do not differentiate between good responses depending on which of the co-primary outcome measures the specified minimum improvement was achieved in, we assume that any person who meets the 'good response' criteria experiences the same utility gain, irrespective of which outcome measure improved for them.

The model used 3-month cycles. *Figure 20* illustrates how participants could move through the model. All participants began in the 'aphasia' health state. Participants could remain in the 'aphasia' health state at 6 months, or transit to the 'good response (6 months)' or 'dead' health states. Participants in the 'good response (6 months)' health state could transit to the 'good response (9 months)' health state, relapse to the 'aphasia' health state or die. Participants in the 'good response (9 months)' health state could transit to the 'good response (12 months and beyond)' health state, relapse to the 'aphasia' health state or die. No new responses could occur beyond 12 months; participants in the 'aphasia' health state at this time point could either remain in that health state or die, whereas participants in the 'good response (12 months and beyond)' health state could remain in that health state, relapse to the 'aphasia' health state or die.

The economic model evaluated the treatment groups included in the Big CACTUS trial, and so transitions between health states included in the model were based on data from Big CACTUS, combined with information on post-stroke mortality rates. For the first 5 years of the model, mortality rates were used from a study of patients who had experienced a stroke ≥ 1 year previously.⁵⁷ After 5 years, additional mortality was applied based on Office for National Statistics life tables,⁵⁸ taking into account the mean age of the Big CACTUS trial participants. We assumed that which health state people resided in did not

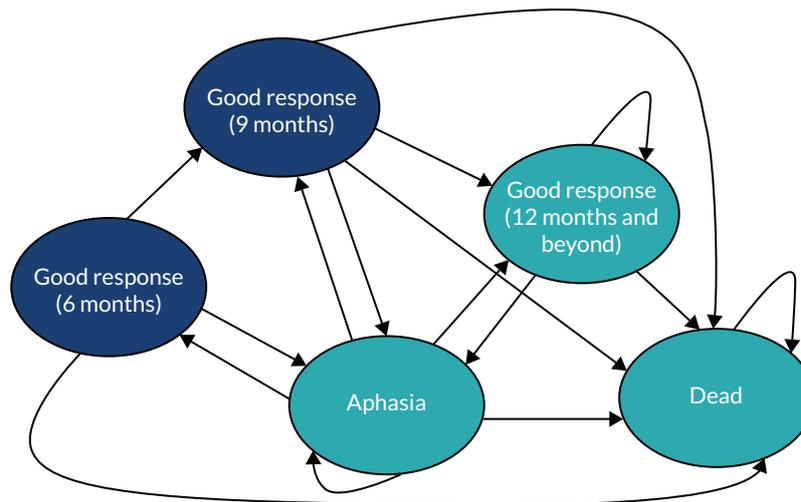


FIGURE 20 Markov model structure. Navy-coloured health states represent tunnel states.

affect the probability of death, and therefore the same probability of death was applied to all health states in the model. This means that we assumed that none of the treatment options under investigation conveyed a survival advantage.

Participants moved into a good response health state if they reported an increase of $\geq 10\%$ in treated words of personal relevance named correctly between baseline and 6 months [or between baseline and 9 or 12 months for the 'good response (9 months)' and 'good response (12 months)' health states, respectively] or if they reported an increase of ≥ 0.5 points since baseline in the activity dimension of the TOMs rating scale). Relapse rates between 6 and 9 months and between 9 and 12 months were estimated by calculating the proportion of participants in the 'Good response (6 months)' and 'Good response (9 months)' health states who did not retain that response at the following time point. Beyond 12 months, the relapse rate was assumed to be equal to that observed between 9 and 12 months.

Health-related quality-of-life utility scores were applied to each health state, and costs were estimated for each of the treatment options under investigation (see *Health-related quality of life and Resource use and costs*). The utility scores applied to each health state were reduced over time to account for ageing according to multipliers estimated by Ara and Brazier,⁵⁹ and QALYs were estimated for each cycle of the model by combining utility scores with life-years, allowing the total QALYs associated with each treatment strategy to be calculated. QALYs were discounted at a rate of 3.5% each year, in line with recommendations made by NICE.⁵²

The model structure requires that we assume that patients who reside within the same health state have the same utility score, regardless of their pathway into that health state. Hence, for example, a new responder at 12 months has the same utility score as someone who responded at 6 months and remained in the good response health state. Similarly, someone who responds and then relapses has the same utility score as someone who never responded. This is a simplification, but modelling further detail would have required additional health states, which would have resulted in estimating utility scores for some health states based on extremely small patient numbers. We believe that our model structure captures the key health states most relevant to the research question.

Health-related quality of life

Big CACTUS trial participants were asked to complete an accessible version of the EQ-5D-5L questionnaire at each of the data collection time points (baseline and 6, 9 and 12 months). An accessible version of the EQ-5D-3L questionnaire was trialled in the CACTUS pilot study.¹⁸ Although accessible versions of the EQ-5D-3L and EQ-5D-5L questionnaires have not been validated, they represent a way in which utility

(HRQoL) scores can be elicited directly from people with aphasia, whose language difficulties may make it difficult to complete standard EQ-5D-5L questionnaires.^{60,61} The accessible version of the EQ-5D-5L used in Big CACTUS is available online (see www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#/) and is described in Whitehurst *et al.*⁶¹ Where applicable, carers of Big CACTUS trial participants were asked to complete a standard EQ-5D-5L questionnaire by proxy at each of the data collection time points. Carers were also asked to complete a standard EQ-5D-5L for themselves at each time point, to investigate potential HRQoL impacts on carers. In line with Al-Janabi *et al.*,⁶² utility scores from the EQ-5D-5L were used to investigate carer utility.

The EQ-5D-5L represents the standard questionnaire used to calculate utilities adopted by NICE.⁵² When the EQ-5D-5L was developed, it was anticipated that this would replace the EQ-5D-3L and hence the five-level version was used in the Big CACTUS trial. Indeed, NICE supports sponsors of prospective clinical studies using the five-level version.^{48,52} However, the English tariff for the EQ-5D-5L published in 2016 is not currently recommended for use in a NICE economic evaluation (but may be recommended in future).^{48,52,63} Instead, NICE currently recommends converting EQ-5D-5L responses to utility scores using the UK cross-walk mapping algorithm developed by van Hout *et al.*⁴⁹ in 2012. This algorithm maps EQ-5D-5L responses onto the EQ-5D-3L tariff in order to generate utility scores. Given the preference for patient-elicited utility scores,⁵² and evidence suggesting suboptimal proxy assessments of utility in the context of stroke,⁶⁴ utility scores calculated using the accessible version of the EQ-5D-5L and the van Hout *et al.*⁴⁹ mapping algorithm were used in our 'base-case' analysis. Several secondary analyses were conducted using:

- the standard EQ-5D-5L questionnaire completed by proxy by carers of participants
- EQ-5D-5L scores calculated using the English EQ-5D-5L tariff,⁴⁷ with no mapping
- an alternative mapping algorithm developed by Hernandez-Alava *et al.*,^{63,65} converting EQ-5D-5L responses to EQ-5D-3L utility scores.

In addition, a secondary analysis was carried out in which QALYs associated with carers were estimated from their self-reported EQ-5D-5L scores and incorporated into the analysis.

In our base-case analysis, MI was used to impute EQ-5D-5L utility scores where data were missing. Predictive mean matching using Stata version 15³⁸ was used to impute missing values of EQ-5D-5L scores at baseline and at the 6-, 9- and 12-month follow-up time points.⁶⁶ An estimate for the missing values was obtained for 10 imputed data sets based on recommendations made in the literature,⁴¹ and Rubin's⁶⁷ framework was used to combine these estimates across imputed data sets to produce an overall averaged estimate. The imputation models were adjusted for variables that were considered to be potentially prognostic for the participant's health status and for which there were full data. These variables were the participant's treatment group assignment (UC/AC/CSLT), the participant's age and the participant's gender. EQ-5D-5L scores for the patients who died during the course of the trial were imputed as zero values after the patient's death. Patients who died were excluded from the analyses estimating the change in utility between responders and non-responders. A secondary 'complete-case' analysis was conducted, which did not impute missing EQ-5D-5L data.

Using the EQ-5D-5L data collected in the trial, we estimated utility scores associated with the 'aphasia', 'good response (6 months)', 'good response (9 months)' and 'good response (12 months)' health states. The utility score associated with the 'aphasia' health state was the mean score across all Big CACTUS participants at baseline. To estimate the impact of achieving a good response on utility scores, we calculated the change in utility score between baseline and the 6-, 9- and 12-month time points for responders and non-responders. By comparing changes in utility score since baseline, we were able to control for differences in baseline utilities that may have existed between participants who subsequently responded and participants who did not.

Resource use and costs

Only costs directly related to the intervention associated with each treatment option were included in the economic evaluation. The pilot CACTUS study collected data on a wide range of resource use (such as medication, primary care and hospital care) but did not show important differences associated with the CSLT intervention with respect to these.¹⁸ This led to the decision not to collect such data in the Big CACTUS trial, and not to include these costs in the economic evaluation. Related to this, it was assumed that the CSLT and AC treatment options were provided in addition to UC, and that UC would not differ between treatment options. This is because we were evaluating therapy delivered by computer in addition to UC received, rather than instead of UC. Given this, the cost over and above UC associated with the UC-only treatment option in the economic model was zero. For further information on the UC actually received during the Big CACTUS trial period, see *Chapter 2, Usual-care control group*. For the CSLT and AC treatment options, the costs associated with delivering these treatments were estimated. Costing was based on the standard approach used in economic evaluations following the three-stage process: identification of resource use, measurement and valuation.⁵⁴

Identification of resource use

The following resources used (costs) were identified for the CSLT treatment option:

- computers/tablets
- the StepByStep computer software
- headset microphones for use with the computer software
- SLT time, related to –
 - initial training
 - senior SLT facilitating training on the CSLT intervention
 - SLTs receiving training from a senior SLT to deliver the CSLT intervention
 - delivery of the intervention
 - SLTs supporting/training SLTAs and volunteers
 - Setting up and personalising/tailoring the computer and software for each participant
 - SLT providing technical support to participants
 - SLT monitoring participant progress
- SLTA time, related to –
 - delivery of the intervention
 - SLTA receiving training/support from a SLT
 - setting up the computer/microphone
 - encouraging the participant
 - assisting with word-finding practice with the software
 - conversations to practise words
- volunteer time, related to –
 - delivery of the intervention
 - volunteer receiving training/support from a SLT
 - setting up the computer/microphone
 - encouraging the participant
 - assisting with word-finding practice with the software
 - conversations to practise words
- travel costs of SLTs, SLTAs and volunteers when visiting the houses of participants to support the intervention

The following costs were identified for the AC group:

- cost of puzzle books
- cost of SLT time spent supporting the AC intervention (a research assistant provided this support during the trial but if the intervention was to be continued in routine clinical practice we assumed that it would be supported by newly qualified SLT or experienced SLTA).

Measurement of resource use

The proportion of CSLT participants who needed to borrow a computer and the proportion who needed headsets were recorded. SLT resource use associated with training on how to use the StepByStep software was based on the number of training workshops held, their length, the average number of SLTs who attended, and time spent by the facilitator, combined with assumptions on how these training sessions would be conducted in reality. For instance, we assumed that in reality training sessions would be held for up to 15 SLTs at a time and that the training would remain beneficial for a 10-year period. This allowed a per-patient cost associated with training to be estimated. Resource use associated with SLTs, SLTAs and volunteers related to the delivery of the intervention and travel costs were measured using activity logs completed by SLTs, SLTAs and volunteers. In these logs, distance travelled for each participant contact was recorded, as was time in minutes per participant related to each of the activity categories listed above.

Valuation of resource use

The Personal Social Services Research Unit (PSSRU) *Unit Costs of Health and Social Care 2017*⁶⁸ were used to place values on time associated with SLT, SLTA and volunteer resource use, with a cost year of 2016/17. We assumed that a SLT at band 7 would facilitate initial training on the StepByStep software, and that band 6 SLTs would mainly deliver the CSLT intervention. SLTAs were assumed to be at band 3, and volunteers were allocated the same unit cost as SLTAs in recognition that they were providing a similar service. SLT costs associated with delivering the AC intervention were costed at band 5. Volunteer costs were included only when the economic evaluation was conducted from a broader perspective. Unit costs associated with each staff type are presented in *Appendix 6*.

The cost of computers for those who required them was based on the average cost of a laptop/tablet purchased through the NHS in Big CACTUS (£690). We assumed that these computers would have a shelf life of approximately 5 years, and that they could be used by 10 people over this time (given the 6-month duration of the CSLT intervention), resulting in a per-participant cost of £69. StepByStep software costs were based on those charged by the manufacturer.²⁰ Travel costs were based on UK Government cost-per-mile estimates for business travel,⁶⁹ and costs associated with the puzzle books used in the AC treatment group were based on the average cost per book incurred in the Big CACTUS trial. All unit costs are presented in *Appendix 6*.

Addressing uncertainty

Distributions were placed around each of the uncertain parameters included in the model (i.e. transition probabilities, utility scores, resource use/costs) for use in probabilistic sensitivity analysis, which allows the uncertainty in all of the modelled parameters to be characterised and permits the estimation of cost-effectiveness acceptability curves (CEACs). The probabilistic sensitivity analysis incorporated 10,000 realisations of the model. Gamma distributions were used for resource use/cost parameters, a beta distribution was used for the 'aphasia' health state utility combined with normal distributions around utility changes associated with a 'good response', and beta distributions were used for probabilities, with dispersions based on numbers observed in the trial. We also undertook expected value of perfect information (EVPI) analysis. The value of information framework allows the maximum value of further research to be estimated, taking into account the uncertainty in the parameters included in the economic model.^{70,71} EVPI was assessed using the Sheffield Accelerated Value of Information tool,⁷² assuming a cost-effectiveness threshold of £20,000 per QALY gained (based on NICE decision rules⁵²)

over a period of 10 years (assuming that it might take 10 years before a new treatment for these patients is developed), using a 3.5% discount rate.

Deterministic sensitivity analyses were undertaken using various alternative techniques for estimating utility scores (see *Health-related quality of life*) and additional scenario analyses were conducted to investigate the impact of different assumptions around the cost of the computer program used by the CSLT intervention. For ease of reference, all secondary analyses are listed in *Table 21*.

Subgroup analysis

Cost-effectiveness was investigated in the prespecified subgroups detailed in *Chapter 3, Subgroup evaluation*:

- Groups with differing severities of word-finding difficulty at baseline, identified using scores from the CAT Naming Objects. Severity was categorised as mild (31–43), moderate (18–30) or severe (5–17).
- Groups with differing impairments in their comprehension ability at baseline as assessed by the CAT Comprehension of Spoken Sentences. Severity was categorised by scores as normal (27–32), mild (18–26), moderate (9–17) or severe (0–8).
- Groups based on the time that had elapsed following the patient's stroke, which were categorised into quantiles ranging from the 25% of patients who entered the trial closest to their stroke to the 25% who had had their stroke longest ago.

TABLE 21 Addressing uncertainty: secondary analyses

Secondary analysis number	Analysis details
1	Complete-case analysis: no imputation for missing EQ-5D-5L data
2	English EQ-5D-5L tariff ⁴⁷ analysis: use tariff directly, with no mapping
3	Carer proxy: using utility scores reported by carers on behalf of patients
4	Alternative EQ-5D-5L mapping algorithm: using Hernandez-Alava <i>et al.</i> ⁶³
5	Carer QALYs: including impact on carer HRQoL
6	All CSLT participants borrow a computer and use a clinician StepByStep licence
7	All CSLT participants use their own computer and an individual StepByStep licence
8	75% of CSLT participants use their own computer and an individual StepByStep licence, 25% borrow a computer and use a clinician licence
9	£120 is paid for a 6-month individual StepByStep licence, rather than £250 for a lifetime licence, reflecting a new payment option
10	Zero cost of StepByStep licence. This is to demonstrate the sensitivity of the cost-effectiveness results to the price of the computer software
11	Zero SLT and SLTA costs. This is to demonstrate the sensitivity of the cost-effectiveness results to costs associated with SLTs and SLTAs
12	Halved SLT and SLTA costs. This is to demonstrate the impact of halving the costs associated with SLTs and SLTAs on the cost-effectiveness results
13	Each intervention group participant received 28 additional hours of face-to-face word-finding therapy from a band 6 SLT instead of 28 additional hours (mean average) delivered by the computer approach. This is to give an idea of the relative cost and cost-effectiveness of the same intervention being provided face to face assuming that it would result in the same outcomes as the CSLT approach for the delivery of word-finding therapy
14	Broader perspective: including costs incurred by volunteers

Methods: within-trial analysis (secondary analysis)

The within-trial economic evaluation was undertaken using individual patient-level data collected within the Big CACTUS trial. The analysis was completed using Stata version 15.³⁸ Utility scores, based on EQ-5D-5L responses, were calculated for participants at baseline and 6, 9 and 12 months. Utility scores for patients who died during the course of the trial were set to zero values after the patient's death. The utility scores at baseline and all follow-up time points were then combined with time for calculating the individual patient-level QALYs using the trapezium rule. As for the model-based analysis, the 'base-case' within-trial analysis used utility scores estimated using the accessible EQ-5D-5L questionnaire and the van Hout *et al.*⁴⁹ mapping algorithm, with MI as described in *Model inputs: health-related quality of life* used for missing values.

The costs included in the within-trial analysis were the same as those included in the model-based analysis, with the exception of the costs associated with training SLTs on how to use the StepByStep software. In the within-trial analysis, this cost reflected exactly what happened in the trial, whereby small numbers of SLTs were trained at a time as soon as research and development permissions were granted for individual sites to begin recruitment. In the model-based analysis, we made slight adjustments to reflect what we would expect to happen in reality (we would expect more SLTs to attend fewer training sessions than provided in the trial because research and development permissions would not be relevant).

Differences between costs and QALYs in the three treatment groups were estimated over the 12-month trial period using seemingly unrelated regression (SUR).⁵⁶ SUR allows for correlation between costs and utility data,^{56,73} and has been used in various trial-based cost-effectiveness analyses.^{74,75} The SUR model was specified to adjust for baseline EQ-5D-5L values as suggested by Manca *et al.*⁵⁵

Uncertainty in the within-trial analysis was explored through the non-parametric technique of bootstrapping to produce 1000 simulations of randomly matched pairs of trial participants (one in the intervention group and one in the control group), from which 1000 ICERs were then calculated.⁶⁶ Cost-effectiveness planes and CEACs were subsequently generated from the 1000 simulations to illustrate the probability of the CSLT intervention being cost-effective. Discounting was not applied to cost and QALY estimates used in the within-trial analysis because the time horizon of the within-trial analysis was 12 months.

Secondary analyses 1–5 and 14 listed in *Table 21* and the subgroup analyses described in *Subgroup analysis* were conducted for the within-trial analysis. Secondary analyses 6–13 listed in *Table 21* were not conducted for the within-trial analysis because these represent hypothetical scenarios that do not represent what happened in the trial period.

Results

The long-term model-based analysis results are presented in *Model-based analysis*. First model input values for transition probabilities, HRQoL and resource use are described based on the Big CACTUS data. Then the model results are presented for the base-case, secondary and subgroup analyses. Results from the within-trial analysis are presented in *Within-trial analysis*.

Model-based analysis

Model inputs: transition probabilities

Transition probabilities between the 'aphasia' health state and the 'good response' health states were based on response rates observed in Big CACTUS. The probability of achieving a good response at 6 months was 0.78 in the CSLT group, compared with 0.46 in the UC group and 0.49 in the AC group.

The probabilities of new responses between 6 and 9 months and between 9 and 12 months were similar between groups (0.33 and 0.39, respectively, for CSLT; 0.37 and 0.37, respectively, for UC; and 0.35 and 0.19, respectively, for AC). Relapse rates between 6 and 9 months and between 9 and 12 months were slightly lower in the CSLT group (0.19 and 0.08, respectively) than in the UC (0.22 and 0.22, respectively) and AC groups (0.28 and 0.19, respectively). Parameter values, distributions and CIs for all transition probabilities included in the model are presented in *Table 22*.

Brønnum-Hansen *et al.*⁵⁷ analysed long-term data on the annual risk of death after stroke. They estimated that after 1 year since stroke the annual risk of death remains approximately constant at 10%. In Big CACTUS, the mean time since stroke was 2.99 years, and hence the 10% risk estimated by Brønnum-Hansen *et al.*⁵⁷ is relevant. This mortality rate was applied to all health states in the model up to year 6, at which point an additional risk of death due to increasing age was added based on mortality data from the Office for National Statistics⁵⁸ and the age and gender split observed in Big CACTUS. This ensured that mortality rates in the model always remained higher than those observed in the age- and gender-matched background population. Notably, Wolfe *et al.*⁷⁶ also present information on long-term outcomes in people who have had a stroke, in a sample collected in London, with people followed up for up to 10 years. The sample size 1 year after stroke was approximately 1000 smaller than that reported by Brønnum-Hansen *et al.*⁵⁷ but results were very similar. The authors provide details of the number of patients followed up annually and the number who died in each 1-year period. These numbers indicate an annual hazard of death of approximately 10% beyond 1 year after stroke. This provides confidence that the mortality rates included in our model are reasonable.

Model inputs: health-related quality of life

The mean baseline utility score across all Big CACTUS participants was 0.61, using MI for missing data, and using the van Hout *et al.*⁴⁹ mapping algorithm. *Appendix 7* shows the proportion of complete data for the EQ-5D-5L questionnaires collected during Big CACTUS. Completion rates were high for the accessible EQ-5D-5L measure and therefore low levels of imputation were required for the base-case analysis.

The change in utility score associated with each response state compared with baseline is presented in *Table 23*, together with the distribution placed around this value in the economic model. Importantly, it can be seen that at 6 and 9 months achieving a good response was associated with a marginal reduction in utility score, whereas at 12 months the difference was positive. All of these estimates were highly uncertain and were not statistically significant.

Similar patterns in utility scores were found when the alternative valuation techniques were used, as shown in *Appendix 8*. Notably, the baseline utility score was significantly lower based on carer proxy questionnaires, but changes in utility associated with response were similarly marginal.

Model inputs: resource use and costs

The cost associated with the CSLT intervention was estimated to be £728.50 in the base-case analysis, compared with £37.70 for AC and £0.00 for UC. Costs in the CSLT group increased to £785.33 when a broader perspective was taken, including costs incurred by volunteers. Details on the breakdown of resource use and cost categories are presented in the following sections.

Hardware and software costs

The majority (66/97, 68%) of CSLT participants needed to borrow a computer to take part in Big CACTUS. In addition, 32 out of 97 (33%) needed headsets to be provided. Hence, hardware costs for the CSLT group were £51.73 per participant (£69 × 0.68 for computers; £14.50 × 0.33 for headsets). AC participants used 4.35 puzzle books on average, resulting in a per-patient hardware cost of £10.89 (£2.50 × 4.35) for the AC group.

We assumed that the 32% of CSLT participants who used their own computers used individual StepByStep software licences, amounting to a per-patient cost of £79.90 (£250 × 0.32). We assumed

TABLE 22 Model parameters: transition probabilities

Parameter	CSLT			UC			AC			Source
	Mean	Distribution (alpha, beta)	Logit 95% CI	Mean	Distribution (alpha, beta)	Logit 95% CI	Mean	Distribution (alpha, beta)	Logit 95% CI	
Probability of good response (0–6 months)	0.78	Beta (73, 21)	0.67 to 0.85	0.46	Beta (45, 52)	0.36 to 0.56	0.49	Beta (49, 40)	0.38 to 0.60	Big CACTUS
Probability of relapse (6–9 months)	0.19	Beta (14, 59)	0.11 to 0.30	0.22	Beta (10, 35)	0.12 to 0.37	0.28	Beta (11, 28)	0.15 to 0.44	Big CACTUS
Probability of new good response (6–9 months)	0.33	Beta (7, 14)	0.15 to 0.57	0.37	Beta (19, 33)	0.24 to 0.51	0.35	Beta (14, 26)	0.21 to 0.51	Big CACTUS
Probability of relapse (9–12 months)	0.08	Beta (5, 61)	0.03 to 0.17	0.22	Beta (12, 42)	0.12 to 0.35	0.19	Beta (8, 34)	0.09 to 0.34	Big CACTUS
Probability of new good response, or renewed response in people who responded at 6 months and relapsed at 9 months (9–12 months)	0.39	Beta (11, 17)	0.22 to 0.59	0.37	Beta (16, 27)	0.23 to 0.53	0.19	Beta (7, 30)	0.08 to 0.36	Big CACTUS
Probability of relapse (12 months onwards)	0.08	Beta (5, 61)	0.03 to 0.17	0.22	Beta (12, 42)	0.12 to 0.35	0.19	Beta (8, 34)	0.09 to 0.34	Big CACTUS
Probability of death (annual)	0.10	Beta (233, 2203)	0.09 to 0.11	0.10	Beta (233, 2203)	0.09 to 0.11	0.10	Beta (233, 2203)	0.09 to 0.11	Brønnum-Hansen <i>et al.</i> (2001) ⁵⁷

TABLE 23 Model parameters: utility scores (accessible EQ-5D-5L)

Parameter	Mean	Distribution (alpha, beta)	95% CI
Utility score (aphasia health state)	0.61	Beta (650.36, 407.99)	0.59 to 0.64
Difference in utility score associated with	Mean	Distribution (standard error)	95% CI
Good response (6 months)	-0.04	Normal (0.03)	-0.09 to 0.01
Good response (9 months)	-0.02	Normal (0.03)	-0.07 to 0.03
Good response (12 months)	0.02	Normal (0.03)	-0.03 to 0.07

that the 68% of CSLT participants who borrowed computers had clinician licences, and assumed that these were bought through the clinician five-licence bundle offered by the StepByStep manufacturer. We assumed that the software had a lifetime of 10 years, and that, although it was feasible that two patients per year could use a licence (for 6 months each), the utilisation rate would be approximately 50%. Thus, 10 patients could benefit from each clinician licence over the lifetime of the product at a per-patient cost of £44 [(£2200/5)/10], amounting to a per-patient cost of £29.94 (£44 × 0.68). The combined per-patient cost of the StepByStep software, taking into account individual and clinician licence costs, was therefore estimated to be £109.84 (£79.90 + 29.94). Hence, we assumed that CSLT participants treated with an individual StepByStep licence would retain that licence for their lifetime (as individual licences cannot be transferred), whereas participants treated with a clinician licence would retain the licence for only 6 months before it is passed on to the next participant, given that the Big CACTUS trial investigated 6 months' usage of CSLT. *Appendix 9* presents details on the computer and software costs associated with scenarios 6 to 10 (for description of scenarios, see *Table 21*). Notably, any scenario that reduces the combined computer plus software cost may also reduce the time for which a patient has access to the computer program, which may affect long-term outcomes.

Uncertainty in hardware and software costs was characterised in the probabilistic sensitivity analysis by placing distributions around proportions of participants who required computers (and therefore those who were treated with individual/clinician licences), the utilisation rate of clinician software licences and the number of puzzle books used (see *Appendix 11*).

Speech and language therapist training costs

Eight 4-hour sessions were held to train 27 SLTs on the use of the StepByStep software. The facilitators estimated that they spent a total of 14 hours preparing for these sessions, with the majority of that time spent preparing for the first session. This amounted to a total of 46 hours of facilitator time and 108 hours of SLT (i.e. trainee) time. Ninety-seven CSLT participants benefited from this training over the 21-month trial period (the recruitment period was 15 months and the intervention period was 6 months). We estimated that the training would remain beneficial for 10 years and that 554 patients [(97 × 12/21) × 10] could therefore benefit from the training over that time period. This resulted in a cost per patient associated with this training of £13.25. In reality, we assumed that only two training sessions would be held to train 27 SLTs, reducing facilitator time to 16 hours (8 hours of training plus 8 hours of preparation). This reduced the per-patient cost estimate to £10.33, which was used in the model-based analysis. In reality, more than 97 participants might benefit from the training in each 21-month period, which would reduce the per-patient cost further. Although we believe that this costing may be conservative, the cost is very low and makes up only a small part of the intervention cost. The uncertainty around speech and language therapy training costs was characterised in the economic model by placing a gamma distribution around the number of patients who could benefit annually from the training (see *Appendix 11*).

Speech and language therapist, speech and language therapy assistant and volunteer time costs

Per-patient time and associated costs incurred by SLTs, SLTAs and volunteers on the tasks described in *Resource use and costs* are presented in *Appendix 10*. Note that costs incurred by volunteers are not included in the base-case analysis. Costs associated with SLTs and SLTAs in the CSLT treatment group totalled £499.22 per patient, more than three times higher than the costs associated with the CSLT computer hardware and software; thus, SLT/SLTA time represents by far the largest cost component of the CSLT intervention. Most SLT time (5.77 hours per patient) was spent setting up the computer software. Costs associated with volunteer time (which are not included in the base-case analysis) were low, totalling £46.22 per CSLT patient. SLT/SLTA costs in the AC group were also low, at £27.29 per patient. Travel costs contributed a further £57.38 to costs incurred by SLTs/SLTAs and £10.61 to costs incurred by volunteers in the CSLT group. Uncertainty around the costs associated with SLT, SLTA and volunteer time was characterised in the economic model by placing distributions around the mean total hours incurred in each of the categories presented in *Appendix 10* (see *Appendix 11*).

Characterising uncertainty

The hardware and software costs, SLT training costs, SLT, SLTA and volunteer time costs and travel costs detailed above are all estimated with uncertainty. *Appendix 11* presents the resource use model parameters that were fitted with distributions to characterise their uncertainty. These were combined with unit costs presented in *Appendix 6* and summed to calculate the costs associated with the CSLT and AC treatment options.

Model inputs: subgroup analyses

Parameter values for transition probabilities, utilities and resource use are presented for all model parameters for the subgroup analyses in *Appendix 12*.

Model results: base case

The model-based long-term cost-effectiveness results (primary analysis) are presented in *Table 24*. The results presented are probabilistic rather than deterministic. This means that they represent the average of 10,000 runs of the model, in which values for the model parameters are randomly drawn from their specified distributions (as defined in *Tables 22 and 23* and *Appendix 11, Table 35*, for transition probabilities, utilities and resource use, respectively). This explains the slight difference in costs estimated compared with the deterministic figures presented in *Model inputs: resource use and costs*. We estimate an incremental cost per patient of £732.73 (95% CrI £674.23 to £798.05) for CSLT compared with UC, and of £694.59 (95% CrI £636.46 to £760.09) for CSLT compared with AC.

TABLE 24 Model-based results: base-case analysis

Analysis	Outcome	Intervention	Comparator	Difference in mean
CSLT vs. UC	Costs (£)	732.73	0.00	732.73 (95% CrI 674.23 to 798.05)
	QALYs	4.2164	4.1992	0.0172 (95% CrI -0.05 to 0.10)
	ICER			£42,686 per QALY gained
CSLT vs. AC	Costs (£)	732.73	38.14	694.59 (95% CrI 636.46 to 760.09)
	QALYs	4.2164	4.1991	0.0173 (95% CrI -0.05 to 0.10)
	ICER			£40,164 per QALY gained
AC vs. UC	Costs (£)	38.14	0.00	38.14 (95% CrI 34.94 to 41.50)
	QALYs	4.1991	4.1992	-0.0001 (95% CrI -0.02 to 0.02)
	ICER			Dominated

We estimated incremental QALY gains per patient of 0.0172 (95% CrI -0.05 to 0.10) for CSLT compared with UC, and of 0.0173 (95% CrI -0.05 to 0.10) for CSLT compared with AC. These result in estimated ICERs of £42,686 per QALY gained for CSLT compared with UC, and of £40,164 for CSLT compared with AC. We estimate that AC is dominated by UC as it is costlier and provides fewer QALYs.

Figures 21–23 depict the cost-effectiveness planes for the comparisons of CSLT vs. UC, CSLT vs. AC and AC vs. UC, respectively. Plots are presented on the same axis scales to aid comparability. These illustrate the results of the model for each of the 10,000 realisations run in the probabilistic analysis, and therefore illustrate the degree of uncertainty associated with the results. The figures show considerable dispersion across the x-axis for all three comparisons, reflecting the high level of uncertainty associated with the estimation of incremental QALYs. It is clearly highly uncertain as to which treatment option is likely to result in the most QALYs gained; for instance, CSLT is estimated to result in more QALYs than UC in 69% of the realisations of the model. There is less uncertainty around costs, demonstrated by a much narrower dispersion across the y-axis.

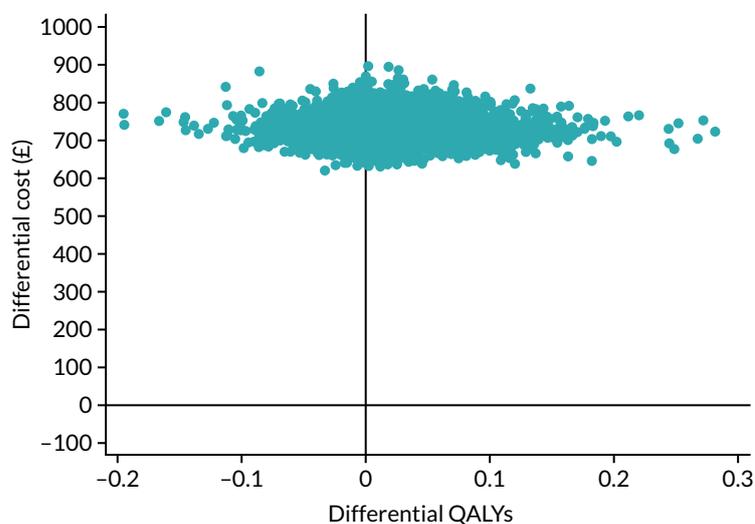


FIGURE 21 Cost-effectiveness plane: base-case analysis – CSLT vs. UC.

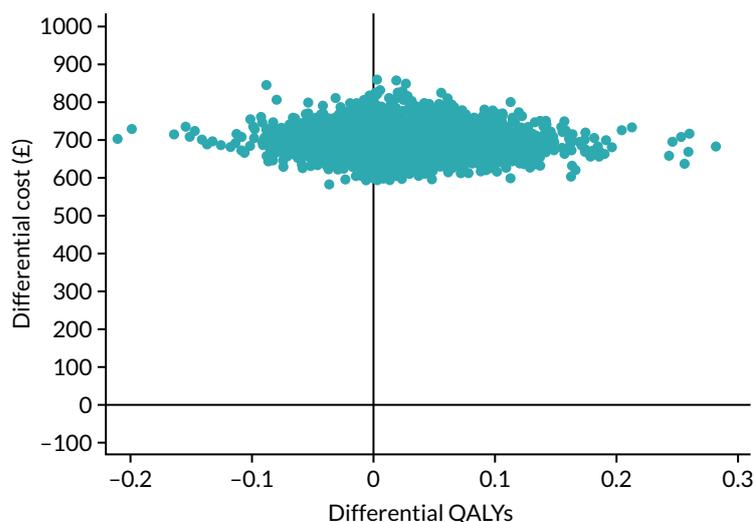


FIGURE 22 Cost-effectiveness plane: base-case analysis – CSLT vs. AC.

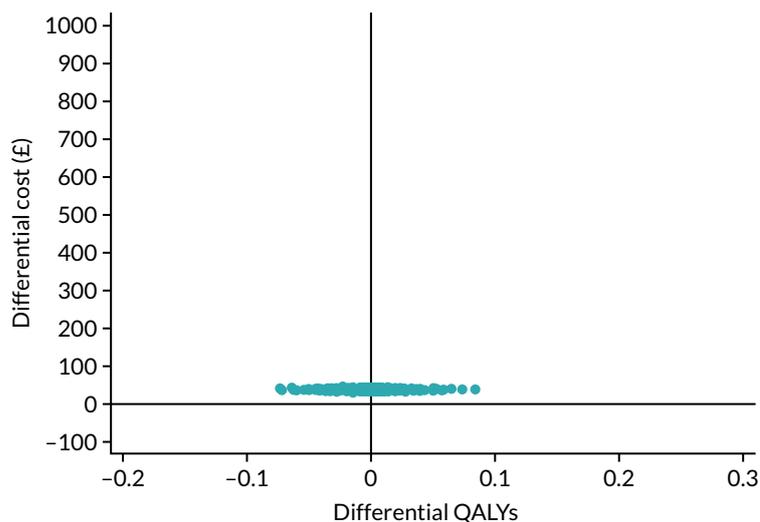


FIGURE 23 Cost-effectiveness plane: base-case analysis – AC vs. UC.

Figure 24 presents the CEAC for the base-case analysis, illustrating the probability of each strategy representing the most cost-effective use of resources for a range of cost-effectiveness thresholds (i.e. the threshold of willingness to pay for an additional QALY). Assuming a cost-effectiveness threshold of £20,000 per QALY gained, the treatment option with the highest probability of representing the most cost-effective strategy is UC, with a probability of 0.56. CSLT and AC both have a probability of 0.22 of representing the most cost-effective strategy at this threshold. At a cost-effectiveness threshold of £30,000 per QALY gained, the probability of UC representing the most cost-effective option is 0.45, compared with 0.32 for CSLT and 0.22 for AC. Hence, although CSLT was estimated to result in the greatest number of QALYs in the majority of the realisations of the model, it provided an ICER of below £20,000–30,000 per QALY gained in a minority of realisations.

The results of the EVPI analysis is presented in *Appendix 13*. We estimated a population-level EVPI of £34.6M. The utility difference associated with a good response at 12 months was the most important parameter in the model, and obtaining perfect information on this parameter was valued at £20.4M at the population level. Hence, if further research was planned, it should be targeted at obtaining further information on this parameter.

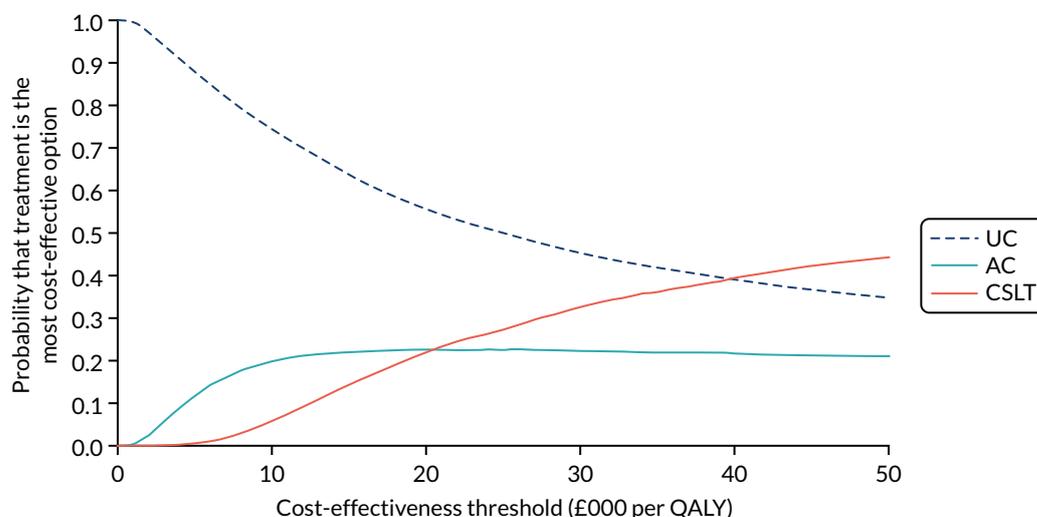


FIGURE 24 Cost-effectiveness acceptability curve: base-case analysis.

Model results: secondary analyses

Appendix 14 presents results from the 14 secondary analyses undertaken (see Table 21). The secondary analyses that investigated alternative costing assumptions (i.e. analyses 6–9) did not result in substantially different estimates of cost-effectiveness. Similarly, conducting the analysis from a broader perspective (including volunteer costs) made little difference, with the ICER increasing slightly to £44,279 per QALY gained for CSLT compared with UC, because volunteer costs were relatively low and affected only the CSLT treatment option. Secondary analyses 10–13 resulted in some more substantial changes in the ICER. Analysis 10 demonstrated that, even if the computer software had zero cost, the ICER would be £36,081 per QALY gained for CSLT compared with UC, because software costs made up a relatively small part of the total intervention cost. In contrast, analysis 11 showed that, if software costs remained the same as in the base-case analysis, but SLT and SLTA costs were zero, the ICER would reduce to £9619 per QALY gained compared with UC. Analysis 12 showed that if software costs remained the same as in the base case, but SLT and SLTA costs were halved, the ICER for CSLT compared with UC would be £26,153 per QALY gained. Analysis 13 aimed to demonstrate the relative cost-effectiveness of the same word-finding intervention delivered face to face rather than via computer software, if the average time that CSLT participants spent doing therapy with the StepByStep computer program (28 hours) was instead delivered face to face by a SLT. Such face-to-face delivery would be appreciably more expensive (with costs approximately double those of the CSLT approach to delivery). Assuming the same effectiveness as the CSLT delivery, face-to-face delivery of additional hours of repetitive word-finding practice would almost certainly not be regarded as cost-effective compared with UC, with an ICER of £78,068.

Cost-effectiveness estimates varied much more substantially in secondary analyses 1–5, which investigated alternative techniques for estimating the QALYs associated with the three treatment options. This further highlights that it is the utility score estimates that represent the key uncertainty in the economic evaluation of the CSLT intervention. Based on a complete-case analysis, the ICER fell to £28,248 per QALY gained, and, importantly, an analysis using utility scores provided by carers by proxy resulted in an ICER of £28,819 per QALY gained for CSLT compared with UC. Using the utility scores provided by carers by proxy, at a cost-effectiveness threshold of £30,000 per QALY gained, the probability of CSLT representing the most cost-effective option was 0.42, compared with 0.39 for UC and 0.19 for AC.

Using the Hernandez-Alava *et al.*⁶³ mapping algorithm instead of that of van Hout *et al.*⁴⁹ resulted in an ICER of approximately £35,000 per QALY gained for CSLT compared with UC. In contrast, directly using the English tariff for the EQ-5D-5L instead of using a mapping algorithm resulted in an increased ICER of £55,639 per QALY gained for CSLT compared with UC. Our base-case analysis represents our preferred analysis because it uses utility scores elicited directly from patients, and uses the NICE-preferred technique for obtaining utility scores from EQ-5D-5L questionnaires. However, all analyses have important weaknesses; for instance, complete-case analyses may be biased owing to selective dropout over time, utility scores reported by proxy may not be representative of the feelings of patients⁶⁰ and the accessible version of the EQ-5D-5L used in our base-case analysis is not validated. Overall, it is clear that the estimated cost-effectiveness of the CSLT intervention depends importantly on how utility scores are estimated.

When we estimated the QALYs associated with the 175 carers who completed the EQ-5D-5L at baseline (analysis 5), we found that CSLT resulted in fewer carer QALYs gained than UC did. This is because carers of participants who achieved a good response at 12 months reported a 0.02 reduction in utility score compared with baseline. This reduction was highly uncertain and is at odds with the 0.01 increase in utility reported by carers of good responders at 9 months. This further highlights the extremely small utility score differences observed in the Big CACTUS trial, and demonstrates that very small positive or negative differences in utility scores (which may have occurred by chance) can substantially change the cost-effectiveness results. We weighted the carer QALYs by a factor of 0.71, because 71% (198/278) of patients had carers at baseline, and then summed patient and carer QALYs to estimate a total QALY gain.

Over the lifetime modelled period, the reduction in carer QALYs was marginally greater than the increase in patient QALYs in the CSLT group compared with the UC group, and hence CSLT was dominated by UC; that is, it was more expensive and provided fewer QALYs. Incorporating carer QALYs could have been done in a variety of ways (e.g. with carer QALYs afforded different weightings), but we did not pursue these further given the finding that including carer QALYs would not improve the ICER for CSLT regardless of the technique used.

Model results: subgroup analyses

Appendix 15 provides cost-effectiveness results associated with the subgroups described in *Subgroup analysis*. Results differed importantly by subgroup, ranging from CSLT being dominated by UC to CSLT being associated with an ICER of < £10,000 per QALY gained. Incremental costs did not change substantially, but QALY increments ranged from positive to negative, resulting in substantially changing ICERs. Again, this highlights the high level of uncertainty associated with the size of the estimated QALY gain associated with CSLT, but also further highlights the uncertainty over the direction of the QALY change; in several subgroups, CSLT resulted in fewer QALYs than UC did.

Perhaps most importantly, the ICER for CSLT compared with UC was < £30,000 per QALY gained for participants with mild (£22,371 per QALY gained) and moderate (£28,898 per QALY gained) word-finding difficulty at baseline, but CSLT was dominated by UC for participants with severe word-finding difficulty at baseline. This was due to slightly larger increases in utility score associated with good responders at 12 months in the mild and moderate word-finding difficulty groups (and also at 9 months for the moderate group) and decreases in utility scores for responders in the severe word-finding difficulty group (see *Appendix 12* for parameter values for subgroups). However, even in these groups, differences in utility scores associated with responders were very small and highly uncertain. The probability of CSLT, UC and AC being cost-effective at a threshold of £30,000 per QALY gained was 0.34, 0.32 and 0.34, respectively, in the mild word-finding subgroup, and was 0.41, 0.37 and 0.22, respectively, in the moderate word-finding subgroup. In the severe word-finding subgroup, these probabilities were 0.09, 0.57 and 0.35, respectively.

It is notable that CSLT is dominated by UC in all comprehension ability subgroups, except the moderate group, for which the ICER is £13,235 per QALY gained. Again, this is driven by whether the estimated utility difference associated with a good response at 12 months is positive or negative. For the moderate comprehension ability subgroup, the estimate utility increase associated with a good response at 12 months is relatively large (0.08), although again this is highly uncertain and, indeed, utility differences associated with a good response at 6 and 9 months were negative in this subgroup.

The subgroup analysis investigating cost-effectiveness according to time since stroke seem to indicate that CSLT represents a cost-effective intervention compared with UC (with ICERs < £20,000 per QALY gained) for all but the quartile of participants who had a stroke most recently. However, further analysis of these subgroups suggests that this is somewhat misleading. For instance, in the third quartile subgroup, the difference in utility score associated with a good response at 12 months was negative; achieving a good response was associated with a reduction in utility. In this subgroup, the response rate was higher for CSLT than for UC, but the relapse rate was also higher for CSLT, leading to fewer CSLT patients than UC patients residing in the 'good response (12 months and beyond)' health state in the long term. This led to CSLT appearing to be beneficial owing to the reduced utility score associated with achieving a good response.

Model results: summary

Our base-case analysis suggests that CSLT is unlikely to represent a cost-effective use of NHS resources compared with UC given a cost-effectiveness threshold of £20,000–30,000 per QALY gained. Incremental costs are relatively low (£732.73), but the incremental QALY difference is also small (0.0172 QALYs), resulting in a base-case ICER of £42,686 per QALY gained. The QALY change is highly uncertain, with CSLT resulting in more QALYs than UC in 69% of the 10,000 realisations of the model run for probabilistic sensitivity analysis. In the base case, the AC treatment option provided

almost exactly the same number of QALYs as UC (-0.0001 incremental QALYs) but was more expensive (although costs were very low, at £38.14 per patient). This resulted in AC being dominated by UC (more expensive and less effective). CSLT produced marginally more QALYs than AC (0.0173-QALY increment) and was associated with an incremental cost of £694.59, resulting in an ICER of £40,164 per QALY gained. At a cost-effectiveness threshold of £30,000 per QALY gained there is a 0.32 probability of CSLT representing the most cost-effective treatment option, a 0.22 probability of AC representing the most cost-effective treatment option and a 0.45 probability of UC representing the most cost-effective treatment option.

Expected value of information analysis demonstrated that the utility difference associated with a good response at 12 months is by far the most important uncertain parameter in the economic model. The magnitude of the change in utility score is important, as is its direction; because the CI associated with the utility change at this time point crosses zero, we cannot be confident that achieving a good response is associated with an improved utility score.

In the base case, achieving a good response was associated with a statistically non-significant decrease in utility at 6 and 9 months, and a statistically non-significant increase in utility at 12 months. The lack of a clear finding around the quality-of-life impact of CSLT is consistent with findings on the co-primary outcome of functional communication ability in conversation and the key secondary outcome of patient perception of their communication and its impact on their life, for which CSLT did not show an effect. The uncertainty around the quality-of-life impact – and the sensitivity of the cost-effectiveness results to this – was further highlighted in secondary analyses in which different techniques for estimating utility scores from the EQ-5D-5L questionnaire resulted in importantly different ICERs for CSLT, ranging from approximately £28,000 per QALY gained to approximately £56,000 per QALY gained compared with UC. Importantly, the ICER was reduced to £28,819 per QALY gained when carer proxy utility scores were used and, in this case, CSLT was the treatment option most likely to represent a cost-effective use of resources at a cost-effectiveness threshold of £30,000 per QALY gained. Subgroup analyses also produced drastically varying results, with CSLT being dominated by UC in some subgroups but producing ICERs < £20,000 per QALY gained in others. Results in the word-finding difficulty subgroups appear more consistent between categories than those found in other subgroup analyses, and we estimate that CSLT is the intervention most likely to represent a cost-effective use of resources in the mild and moderate word-finding difficulty subgroups, given a cost-effectiveness threshold of £30,000 per QALY gained.

The CrI around the incremental QALY gain associated with CSLT is reasonably narrow, but it crosses zero. Given its low cost, only a small QALY gain is required in order for CSLT to appear cost-effective, and very small changes in QALY estimates can have large impacts on the ICER. However, we are uncertain about whether the intervention leads to increased QALYs or not. Furthermore, when QALY increments are close to zero it may be expected that some subgroups will result in point estimates of QALY reductions whereas others will result in point estimates of QALY increases. Overall, we are highly uncertain about whether the CSLT intervention provides more or fewer QALYs than UC and AC.

In contrast, the value of information analysis demonstrated that the uncertainty around cost parameters included in the model was not valuable and secondary analyses that investigated alternative software and computer costs had little impact on the ICER. However, these analyses also demonstrated that the cost-effectiveness results were sensitive to the most important cost driver, which was SLT time. If SLT and SLTA costs could be halved, the ICER for CSLT compared with UC would be £26,153 per QALY gained. Most SLT time was spent setting up the computer program, and it may be feasible that cost savings could be made. For instance, if it were possible for SLTs to set up the computer software in 1 hour, instead of the average 5.77 hours observed in Big CACTUS, SLT and SLTA costs would be reduced by approximately 38%, resulting in an ICER of £30,181 per QALY gained.

Within-trial analysis

Inputs: health-related quality of life

The advantage of a within-trial analysis is that the uncertainty associated with extrapolating beyond the trial period is avoided. The disadvantage is that this results in an evaluation of only the short-term cost-effectiveness of the interventions, with longer-term impacts being ignored.

For the within-trial analysis, a model was not needed; QALYs were simply estimated for each treatment option according to EQ-5D-5L responses reported during the trial, adjusted for any baseline differences between treatment groups using SUR.

Figure 25 illustrates the trend of mean accessible EQ-5D-5L scores with 95% CIs at baseline and the follow-up points (6, 9 and 12 months) for the UC, AC and CSLT treatment groups, where missing scores were imputed as described in *Health-related quality of life*, and where utility scores were derived using the van Hout *et al.*⁴⁹ algorithm. Similar figures are presented in *Appendix 16* for a 'complete-case' analysis (where missing values were not imputed) and for an analysis where only data from participants who provided responses at all follow-up time points were included. *Appendix 16* also includes these results in tabular form, and presents similar results for proxy EQ-5D-5L scores and carer EQ-5D-5L scores.

Results suggest that imputation had a small impact on estimated utility scores, although differences were slightly larger when comparing the utility scores based on imputed data with those estimated including only patients who provided complete responses at all time points. This is not unexpected, because most data are lost when only patients who provide complete responses at all time points are included. Importantly, changes in utility over time are extremely marginal for all treatment groups. However, it appears that utility scores decline slightly in the CSLT group whereas they increase slightly in the UC group.

Inputs: resource use and costs

Costs associated with the three treatment options were identical to those included in the modelled analysis (see *Model inputs: resource use and costs*), with the exception of the cost associated with training on the StepByStep software for SLTs. For the model-based analysis, we made an assumption that, in practice, training sessions larger than those in the Big CACTUS trial would be held. For the within-trial analysis, we instead based this cost on exactly what was observed during the trial. This makes very little difference: the per-patient cost associated with this training was £10.33 for the model-based analysis, and is £13.25 in the within-trial analysis (see *Model inputs: resource use and costs*). The mean incremental cost of CSLT compared with UC was £732.37 (£789.57 from the broader perspective) across the 1000 bootstrapped simulations of randomly matched pairs of trial participants undertaken for the within-trial analysis.

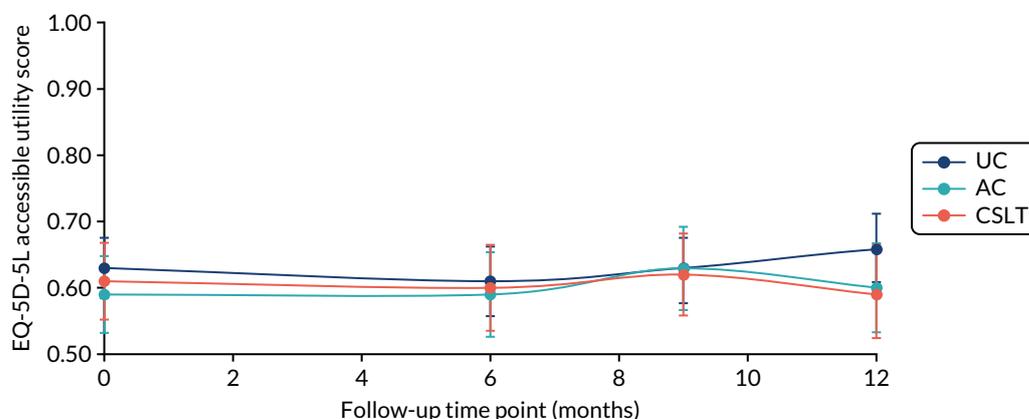


FIGURE 25 Means and 95% CIs of accessible EQ-5D-5L scores by treatment group at each time point after missing EQ-5D-5L scores are imputed.

The mean incremental cost of CSLT compared with AC was £694.65 (£751.85 from the broader perspective). *Appendix 17* presents histograms for the distributions of costs in the CSLT and the AC groups, respectively. In the CSLT group, costs were positively skewed, demonstrating that the vast majority of CSLT participants incurred costs of between approximately £250 and £1000, but a small number of participants incurred costs of > £1500. In the AC group, costs follow a more normal distribution. These costs were much less prone to variation between participants owing to the much smaller role of SLTs.

Results: base case

Table 25 presents the results of the SUR used to conduct the within-trial analysis, complete with non-parametric bootstrapping to characterise uncertainty. QALYs during the trial period were estimated to be higher in the UC and AC treatment groups than in the CSLT treatment group, but this was highly uncertain, with CIs crossing zero. The incremental QALYs of CSLT compared with UC were estimated to be -0.007 ($p = 0.71$). The incremental QALYs of CSLT compared with AC were estimated at -0.007 ($p = 0.73$). CSLT led to an increase in costs compared with UC (£732.37, $p < 0.001$) and AC (£694.65, $p < 0.001$) and therefore CSLT was dominated by UC and AC in the within-trial base-case analysis. AC was estimated to result in fewer QALYs during the trial period than UC (-0.004 increment, $p = 0.83$) and was more expensive; therefore, UC dominated AC.

The cost-effectiveness plane for the within-trial base-case analysis comparing CSLT with UC is presented in *Appendix 18*. The plane shows considerable uncertainty in relation to estimation of incremental QALYs with a wide scattering of points across the x-axis. Approximately 36% of the estimates resulted in a QALY gain for CSLT, and approximately 64% resulted in a QALY gain for UC.

The cost-effectiveness plane for the within-trial base-case analysis comparing CSLT with AC is presented in *Appendix 18*. The plots are very similar to those for the comparison of CSLT with UC; there is a high degree of uncertainty around incremental QALY estimates, with approximately 38% of the estimates resulting in a QALY gain for CSLT and approximately 62% resulting in a QALY gain for AC.

The cost-effectiveness plane for the within-trial base-case analysis comparing AC with UC are presented in *Appendix 18*. Again, there is a high level of uncertainty as to which treatment option will result in the most QALYs. Approximately 42% of the estimates resulted in a QALY gain for AC and approximately 58% resulted in a QALY gain for UC.

TABLE 25 Within-trial analysis results: base case

Analysis	Outcome	Intervention, mean (SD)	Comparator, mean (SD)	Difference in mean (95% CI)
CSLT vs. UC	Costs (£)	732.37 (344.22)	0.00 (0.00)	732.37 (665.59 to 799.15)
	QALYs	0.6058 (0.23)	0.6259 (0.19)	-0.007 ^a (-0.05 to 0.03)
	ICER			Dominated
CSLT vs. AC	Costs (£)	732.37 (344.22)	37.72 (16.02)	694.65 (619.54 to 769.75)
	QALYs	0.6058 (0.23)	0.5987 (0.21)	-0.007 ^a (-0.05 to 0.03)
	ICER			Dominated
AC vs. UC	Costs (£)	37.72 (16.02)	0.00 (0.00)	37.72 (34.61 to 40.82)
	QALYs	0.5987 (0.21)	0.6259 (0.19)	-0.004 ^a (-0.04 to 0.04)
	ICER			Dominated

^a The difference in mean QALYs was adjusted for baseline EQ-5D-5L utility scores.

The CEAC for the within-trial base-case analysis is presented in *Appendix 18*. The CEAC illustrates the probability that each strategy is the most cost-effective option for a range of cost-effectiveness thresholds (£0–50,000 per QALY). At a cost-effectiveness threshold of £20,000 per QALY gained, the treatment option with the highest probability of representing the most cost-effective strategy is UC, with a probability of 0.5. CSLT has a probability of 0.15 and AC has a probability of 0.35 of representing the most cost-effective strategy at this threshold. At a cost-effectiveness threshold of £30,000 per QALY gained, the probability of UC representing the most cost-effective option is 0.48, compared with 0.18 for CSLT and 0.34 for AC.

Results: secondary analyses

The results from the six secondary analyses described in *Table 21* are presented in *Appendix 19*. Note that results for analysis 14 are not included for the comparison of AC and UC because volunteer costs were not incurred in the AC and UC groups.

Results of the secondary analyses demonstrate that the technique used to estimate utilities from the EQ-5D-5L questionnaire is very important; depending on the technique used, CSLT either produces QALY gains or QALY losses compared with UC. QALY losses are estimated for the base-case analysis (using the van Hout *et al.*⁴⁹ algorithm to estimate utility scores), when the English EQ-5D-5L tariff is used (with no mapping algorithm) and when the Hernandez-Alava *et al.*⁶³ mapping algorithm is used. In contrast, QALY gains are estimated for CSLT compared with UC when carer proxy EQ-5D-5L utility scores are used and when complete data are used at each time point (with no imputation for missing values). However, even when QALY gains are estimated, the ICER for CSLT compared with UC remains > £50,000 per QALY gained.

When the utility scores associated with the 175 carers who completed the EQ-5D-5L at baseline were incorporated into our analysis (analysis 5), the ICER for CSLT compared with UC reduced to approximately £127,000 per QALY gained, and for CSLT compared with AC reduced to approximately £81,000 per QALY gained. This is in contrast to the results of the model-based analysis when carer utilities were incorporated, which resulted in CSLT being dominated by UC. In the model-based analysis, utility scores reported at 12 months are most important, because these are extrapolated into the future. In the within-trial analysis, the HRQoL experienced during the trial period is of most importance, with the 12-month utility score playing a relatively small part. This suggests that during the trial period carers of CSLT participants fared better than carers of UC participants. However, again, incremental QALY estimates were highly uncertain in this scenario, with CIs crossing zero.

As for the model-based analysis, adopting a broader perspective (including volunteer costs) has very little impact on the within-trial cost-effectiveness analysis results. The incremental cost associated with CSLT is marginally increased compared with UC and AC, with no difference made to QALYs, and thus CSLT remains dominated by both other treatment options.

The sensitivity of results to the technique used to calculate utility scores from the EQ-5D-5L responses is further highlighted by the results comparing AC with UC (see *Appendix 19*). Although AC was dominated in the base-case within-trial analysis, it is associated with ICERs of approximately \leq £10,000 per QALY gained using all alternative valuation techniques. Differences between QALYs estimated for AC and UC are even smaller than for CSLT compared with UC, which explains why the results of comparisons between AC and UC are even more sensitive to the technique used to derive utility scores than for the CSLT and UC comparison.

Results: subgroup analyses

Appendix 20 reports within-trial analysis cost-effectiveness results for the subgroups described in *Subgroup analysis*. Again, results differed importantly by subgroup, ranging from CSLT being dominated by UC to CSLT being associated with an ICER of < £30,000 per QALY gained. Similar fluctuations in results were observed for the comparison with AC and for the comparison between AC and UC. As for

the model-based analyses and the secondary within-trial analyses, incremental costs did not change substantially between groups, but QALY increments ranged from positive to negative, resulting in substantially changing ICERs. Again, this highlights the high level of uncertainty associated with both the size and direction of the estimated QALY gain or loss associated with CSLT.

Few consistent patterns can be observed when comparing the results across subgroups. However, it is worthy of note that CSLT is dominated by UC in the mild and moderate word-finding difficulty subgroups, but has an ICER of < £30,000 per QALY gained in the severe subgroup. This is in direct contrast to the results produced by the model-based analyses.

Summary

Our base-case within-trial analysis suggests that CSLT is unlikely to represent a cost-effective use of health-care resources compared with UC and AC treatment options. Over the trial period, UC and AC were both estimated to result in marginally more QALYs than CSLT, while costing less. The number of QALYs associated with each of the treatment options was very similar and differences between the treatment options were highly uncertain. Secondary analyses demonstrated that using different techniques to derive utility scores from the EQ-5D-5L responses could change the direction of the estimated differences for all comparisons. This emphasises the very small magnitude in the difference between QALYs gained associated with the different treatment options. Results of subgroup analyses were also extremely variable, making it very difficult to draw conclusions on cost-effectiveness.

It is notable that the base-case results of the within-trial analysis are in direct contrast to the base-case results of the model-based analysis. This is due to the shorter time frame considered in the within-trial analysis. At the 6- and 9-month time points, good response was associated with very small reductions in utility compared with no response; it was only at the 12-month time point that this difference became positive. As a result, if the time frame of the base-case model-based analysis is set to 12 months, it also results in CSLT producing fewer QALYs than UC and AC. It is the positive difference in the utility score associated with good response at 12 months that leads to the estimated increase in QALYs associated with CSLT compared with UC and AC in the model-based analysis. This utility score is extrapolated into the future and because CSLT results in more participants residing in the 'Good response' health state in the long term, QALY gains are accrued. The within-trial analysis is restricted to a 12-month time frame and therefore does not project this gain into the future; hence, potential future QALY gains are not accrued.

The health economic analysis is further discussed in *Chapter 6*.

Chapter 6 Discussion and conclusions

Summary and interpretation of clinical findings

The co-primary outcome measures in this trial were chosen to establish whether or not self-managed CSLT tailored by SLTs and supported by volunteers or assistants for word-finding:

- increases the ability of people with aphasia to retrieve vocabulary of personal importance
- improves functional communication ability in conversation.

Impact on word-finding

The improvement in retrieval of treated words at 6 months post randomisation was substantially greater with CSLT than with UC or AC. On average, CSLT improved word-finding by 16.2% more than UC (95% CI 12.7% to 19.6%; $p < 0.0001$) and 14.4% more than AC (95% CI 10.8% to 18.1%), showing that the effect was largely due to speech and language therapy components of the CSLT intervention and not just extra activity and attention. The effect was in excess of the prespecified minimal clinically important difference of 10%. This effect was mostly maintained at 9 and 12 months, suggesting maintenance of the treatment effect following the 6-month intervention period. It should be noted that 57 (61%) of the participants in the CSLT group continued to use the computer therapy unsupported beyond 6 months, which may have supported this maintenance. Of the 57 who practised beyond 6 months, 33 practised between 6 and 9 months and 24 practised beyond 9 months.

Functional communication, patient perception and quality of life

There was no significant improvement in the co-primary outcome measure of functional communication in conversation in any of the groups, providing no evidence that the CSLT approach we evaluated for word-finding improves communication in conversation. Our key secondary outcome measure suggests that CSLT did not result in improvement in participants' own perceptions of their communication and participation and the impact of these on their life, which is unsurprising given that no improvement in functional communication was seen. These findings indicate that CSLT supported participants with aphasia to make significant improvements in their ability to retrieve words of personal importance at the impairment level, but these gains did not generalise to functional communication settings, thus limiting the impact of the improvements in their lives.

Generalisation of improvement in word-finding to conversation was further explored through a secondary outcome measure specifically identifying the difference in the number of treated words used in conversations structured around topics that provided the opportunity to use those words. In keeping with the co-primary outcome measure of conversation, the groups were similar with respect to the number of words used in conversation. However, a post hoc analysis showed that 1 in 10 participants in the AC and UC groups, and 3 in 10 participants in the CSLT group used at least five more treated words in conversation at 6 months than at baseline, suggesting that there may be a small treatment effect of the use of treated words in the functional context of conversation in a few participants. The lack of carryover of improvement from retrieval of treated words in a confrontation naming task to use in functional communication situations may be a result of people being used to communicating in a different way (e.g. gesture) and therefore not remembering to use new words when they have the opportunity. In addition, retrieving words in functional communication settings is likely to be a subtly different to or a more difficult task than naming a picture. Further intervention may be required to help people with aphasia use their 'new' words in useful situations.

Generalisation to untreated words

A further secondary research objective of the trial was to investigate whether or not learning of treated words generalises to being able to retrieve words that were not specifically treated. There was no significant difference between the three trial groups in retrieval of untreated words between baseline and 6 months. Therefore, improved word-finding of treated words did not result in generalisation to untreated words. This suggests that people get better at saying what they practise saying > 4 months post stroke. Because individuals with long-term aphasia may learn only the words they practice, it is most useful to them if these words are of personal relevance.

Impact of aphasia profile, word-finding severity and length of time post stroke

Prespecified subgroup analyses indicated that the effect of CSLT on word-finding was slightly higher for participants with mild word-finding difficulties and for those whose verbal comprehension was within normal limits. It is likely that the brain lesions of these participants were smaller, allowing greater retained potential to relearn words with practice. Importantly, subgroup analyses showed that the treatment effect was broadly consistent regardless of the time post stroke, suggesting that people with aphasia may learn new words at any time after stroke (range of 4 months to 36 years in our trial).

Carer perception of communication effectiveness and impact on their quality of life

The carers also rated their perception of the participant's communication effectiveness and their own quality of life. For communication effectiveness, there was a MDC of 4.6% (95% CI 0.3% to 9.0%) in favour of CSLT compared with UC at 6 months and a MDC of 5.1% (95% CI 0.5% to 9.7%) in favour of CSLT compared with AC, indicating a small improvement in the carers' perception of communication effectiveness with CSLT. However, the long-term effects of the intervention on the average change in the carer-rated communication effectiveness were very small: 0.6% (95% CI -4.4% to 5.7%) and 2.7% (95% CI -1.9% to 7.4%) in favour of CSLT compared with UC at 9 and 12 months, respectively. This indicates that any perception of improvement in communication effectiveness is unlikely to be maintained long term. There was also a 5.3% (95% CI -1.1% to 11.7%) MDC in perception of carers' quality of life in favour of CSLT compared with UC at 6 months. However, the improvement in carers' reported quality of life in the CSLT group compared with AC was close to zero (0.3%, 95% CI -6.4% to 6.9%). This suggests that the small improvements seen in carers' quality of life at 6 months may be due to the increased levels of activity their relative is engaged in and the receipt of increased amounts of attention in the CSLT and AC interventions compared with UC alone.

Safety

We investigated the safety of the CSLT intervention. Negative effects of CSLT were low, with 27% of participants in the CSLT group reporting fatigue or anxiety at some point, which translates to an average of only one event per person per year. Effects on eyes or headaches were very rarely reported. On average, the incidence of AEs per participant per person-year of follow-up was 2.18, 1.79 and 1.87 in the CSLT, AC and UC groups, respectively. This indicates a slight increase in all AEs in the CSLT group, with an IR of 1.16 (95% CI 0.83 to 1.62) and 1.22 (95% CI 0.85 to 1.77) compared with the UC and AC groups, respectively, although we cannot rule out similarity in incidences between groups as the CIs include an IRR of 1. Fits and visual difficulties were uncommon, but the incidence of fits was reported to be three times higher in the CSLT group than in the UC or AC groups, although such events were rare in all groups. Similarly, slightly more AEs were reported in the CSLT group than in the UC or AC groups; however, the CSLT group had more opportunity for reporting AEs as participants were prompted with a negative effects form each month in addition to the 3-monthly check by outcome assessors received by all groups, whereas participants in the AC and UC groups were not. Differences in the incidence of AEs between groups were insufficient to suggest differences in risk levels. The number of SAEs was 18 (15.8%) in the UC group, 11 (13.9%) in the AC group and 9 (10.6%) in the CSLT group. Although there were fewer SAEs in the CSLT group, there is insufficient evidence to suggest differences in IRs of SAEs across interventions.

Summary and interpretation of health economic findings

In this trial, we sought to establish whether or not CSLT is cost-effective for persistent aphasia post stroke. Unfortunately, estimates of the quality-of-life benefit associated with the intervention were unstable and highly uncertain. Although the CrI around the incremental QALY gain associated with CSLT was reasonably narrow, it crosses zero, and we are uncertain about whether or not the intervention leads to increased QALYs, making it very difficult to make firm conclusions on its cost-effectiveness. A model-based analysis in which costs and QALYs were extended over a lifetime period formed the primary analysis. A within-trial analysis was also conducted in which the costs and QALYs were measured over the 12-month trial period and were not extended further.

Main findings

The base-case (primary) model-based cost-effectiveness analysis resulted in an ICER of £42,686 per QALY gained for CSLT compared with UC, and an ICER of £40,164 per QALY gained for CSLT compared with AC. AC was dominated by UC, meaning that the UC group cost less and had greater quality-of-life gains (measured on EQ-5D-5L for the health economic analysis) than the AC group. These ICERs are higher than the current NICE cost-effectiveness threshold of £20,000–30,000 per QALY gained,⁵² suggesting that in the UK the CSLT approach we evaluated is unlikely to be considered to represent a cost-effective use of health-care resources for the whole population with word-finding difficulties as a result of post-stroke aphasia.

However, cost-effectiveness estimates were highly uncertain. Although results were robust to altering assumptions around software and computer costs, they were very sensitive to the utility score estimates associated with achieving a good response to treatment, which were highly uncertain. It is not clear whether or not the intervention leads to an increase in quality of life as measured by the EQ-5D-5L. The cost-effectiveness results were sensitive to using different techniques to derive utility scores (a measure of HRQoL) from EQ-5D-5L responses. The base-case (primary) analysis used the EQ-5D-5L mapping algorithm developed by van Hout *et al.*⁴⁹ to calculate utility scores, as recommended in NICE guidelines for cost-utility analysis.⁵² However, an unvalidated accessible version of the EQ-5D-5L⁶¹ was used so that utility scores could be derived directly from trial participants. In circumstances in which trial participants cannot complete a standard EQ-5D-5L questionnaire themselves, a typical approach is to use scores derived from standard EQ-5D-5L questionnaires completed by carers by proxy. When we conducted secondary analyses using this technique, the ICER for CSLT compared with UC fell to £28,819, potentially representing a cost-effective use of health-care resources. Using an alternative mapping algorithm (that developed by Hernandez-Alava *et al.*⁶³) to derive utility scores from the accessible version of the EQ-5D-5L also led to a reduced ICER for CSLT compared with UC, equal to £34,921 per QALY gained. These results are consistent with findings from work by the NICE Decision Support Unit comparing the mapping methods.⁶³ In contrast, directly using the English tariff for converting EQ-5D-5L responses to utility scores (with no mapping algorithm) led to an increase in the ICER, to £55,639 per QALY gained.

In addition, we found that important reductions in the ICER could be achieved if SLT costs could be reduced substantially. For instance, if the computer software could be set up for participants by SLTs in 1 hour instead of the average 5.77 hours observed in the Big CACTUS trial, the ICER for CSLT compared with UC would decrease to £30,181 per QALY gained. However, this ICER would remain highly uncertain owing to the uncertainty surrounding the QALY gain. In the trial, we did not evaluate the impact of the involvement of SLTs in the intervention. However, as SLT tailoring of the software incurred a significant proportion of the intervention cost, it is important to understand if and how this is related to the clinical outcomes. The relationship between SLT support and intervention effectiveness is the focus of a completed PhD.

Explanation of findings

The sensitivities on the ICERs according to the method used to derive utility scores demonstrate that the key uncertainty in this economic evaluation surrounds the size and direction of the QALY gain or loss associated with CSLT compared with UC and AC. A key message is that CSLT did lead to significantly more participants achieving a 'good response' in naming of words with personal relevance or conversation than UC and AC did. However, the impact on utility of achieving a good response appeared to be small and highly uncertain. Given the low cost associated with CSLT, very small increases or decreases in the estimated QALY gain can have a large impact on the ICER. In the vast majority of analyses completed (primary, scenario and subgroup analyses), the CIs for the QALY gain associated with CSLT compared with UC ranged from negative to positive. A related issue surrounds the definition of a 'good response'. Although CSLT led to substantially more participants achieving a 'good response', large proportions of participants in the AC and UC groups also achieved a 'good response'. Potentially, the definition of response (improvement of 10% in word-finding or an improvement of 0.5 on the TOMs) may have been too lenient; differences in quality of life may be more likely in participants who achieve a greater response. Related to this, an alternative modelling approach could have split the response state according to whether a response was achieved through a word-finding improvement or an improvement on TOMs (or both). In practice, as would be expected given the clinical results, the majority of responses were due to improvements in word-finding; for instance, at 6 months only 8 of the 73 responders in the CSLT group achieved a response based on TOMs and not on word-finding, and 23 achieved a response on both measures. Hence, splitting the response category would have involved estimating utility scores for different response groups based on very small patient numbers and would have been prone to substantial error.

In the base case (the primary health economic analysis), the utility score in patients who achieve a good response is estimated to be worse at 6 and 9 months than that in patients who do not achieve a good response. At 12 months, good responders are estimated to have marginally higher utility than non-responders. Given that all CIs overlap, these alternative directions in utility differences may be due to chance. Alternatively, they may suggest that it takes time for a good response to lead to utility gains as measured by the EQ-5D-5L.

Given the clinical results observed in Big CACTUS, it may not be surprising that CSLT had little impact on utility scores. Although the intervention had a significant impact on word-finding ability, it had no discernible impact on functional communication measured or on participants' perceptions of communication effectiveness or impact on their lives. Therefore, it seems unlikely that a response would have an impact on the EQ-5D-5L domains to indicate quality of life, and indeed our within-trial analysis estimated marginally more QALYs for the UC and AC treatment groups than for the CSLT group, although this was highly uncertain. If the improvement in word-finding ability could be converted into a functional improvement, it is possible that QALY gains could be derived and, given the low cost of the intervention, this could result in CSLT being deemed cost-effective. Research to investigate this further may be valuable. In our value of information analysis, we found that the expected value of obtaining perfect information on the change in EQ-5D-5L score associated with a good response at 12 months is £20.4M. This suggests that it would be highly valuable to collect more data on this model parameter to enable it to be quantified more accurately.

Subgroup analysis

Importantly, the prespecified subgroup analyses (of the model-based health economic evaluation) suggested that the CSLT approach we evaluated may represent a cost-effective use of health-care resources for participants with mild or moderate word-finding difficulty. In these groups, CSLT was associated with ICERs of £22,371 and £28,898 per QALY gained, respectively, compared with UC, and of £30,911 and £18,855 per QALY gained compared with AC. In contrast, CSLT was associated with a higher cost and lower quality-of-life scores than AC and UC for participants with severe word-finding difficulty (CSLT was dominated by AC and UC). The proportion of CSLT participants who achieved a good response was similar in the different word-finding difficulty subgroups, but cost-effectiveness was improved in the mild and moderate groups owing to a slightly increased utility difference associated

with a good response at 12 months in these groups, and a negative difference (compared with no response) in the severe subgroup. However, these analyses remain highly uncertain and, as for the base-case analysis, achieving a good response was associated with a reduction in utility at earlier time points for both the mild subgroup and the moderate subgroup. Again, the direction of the utility changes may have been due to chance. In addition, within-trial analyses in the same subgroups provided opposite results (whereby CSLT was dominated by UC in the mild and moderate word-finding difficulty subgroups). Related to this, it is important to reiterate that cost-effectiveness results based on within-trial and model-based analyses were often substantially different because the estimated higher EQ-5D-5L score associated with CSLT at the 12-month time point was extrapolated for the remaining lifetime period of the model-based analysis, but was not extrapolated in the within-trial analysis.

Cost of computerised speech and language therapy intervention

The intervention cost £733 per participant, which is relatively low. This enabled a mean of 28 hours of independent, repetitive word-finding practice. Although we do not know the relative effectiveness of 28 hours of therapy provided face to face by a SLT, providing an additional 28 hours of face-to-face speech and language therapy by a mid-grade (Agenda for Change band 6) SLT would cost £1400, almost twice as much as supporting an individual to practise independently with a computer. If the time taken by SLTs to set up the computer program could be reduced, the cost-effectiveness of the intervention could improve markedly.

Fidelity to the interventions

Intervention coverage was excellent as all participants received a computer with the software on it.

Quality of computerised speech and language therapy intervention delivery

Measures used as indicators of quality of intervention delivery included training, intervention knowledge of providers, appropriate tailoring of the computer therapy and provision of support to the participants. There was high/good fidelity to these measures of intervention delivery quality. Regarding support provided by volunteers/SLTAs, 85% of the participants received support for using their newly learned words in conversation or functional contexts; however, these skills were practised with the volunteer or SLTA for a total of only 45 minutes (median) per participant across the 6-month intervention period. Increased amounts of time spent on these transfer activities may be required to assist with greater use of the newly learned words in conversation. Only 52% of feedback forms from volunteers/SLTAs were monitored by SLTs, indicating lower adherence to the role of monitoring the SLTA/volunteer support provided to the participant.

Adherence to computerised speech and language therapy practice

The CSLT participants practised for a mean of 28 hours, just above the recommended minimum for high adherence (26 hours); the median was below this at 21 hours, with just fewer than half of the participants (46%) meeting this minimum for high adherence. Sixty-four per cent of participants practised for a minimum of 10 hours, indicating some adherence to the intervention. Although we are able to report the amount of practice time, further analysis has been conducted during a PhD study regarding the content of the practice and motivation to practise.

Conclusion for computerised speech and language therapy

Overall fidelity to the CSLT intervention was judged to be high to fair.

Fidelity to attention control

The AC intervention was to complete a puzzle each day with a supportive telephone call from the research team each month for 6 months. As we expected periods of illness and holiday as with the CSLT group, we expected a minimum of four telephone calls to have taken place and four to six puzzle books to be completed. It was difficult to measure the adherence to puzzle book practice. Only 18% of

participants had six books and four telephone calls but 61% of participants had a minimum of four puzzle books and four telephone calls, suggesting that adherence to the AC group was fair and therefore similar or slightly lower than the adherence to the CSLT intervention.

Fidelity to usual care

The UC received during the trial was generally similar between groups, suggesting that the adherence to the provision of UC was consistent across groups. During the 6-month intervention period, the mean amount of UC received by all participants in the UC, AC and CSLT groups, respectively, was 3.8 hours, 3.2 hours and 3.2 hours. The amounts of UC provided decreased over the trial period in all groups. This is likely to be because participants progressed along the stroke pathway of care as the trial progressed. The pre-baseline UC data indicated that patients receive reduced amounts of UC with time post stroke.

Trial results in the context of other studies

Age and gender

The age and gender profile of the participants of the Big CACTUS trial was in keeping with that seen in other aphasia studies (not all RCTs, therefore open to selective sampling), 60.7% male being consistent with 60.8% of the Rehabilitation and recovery of people with aphasia after stroke (RELEASE) data set⁷⁷ of 5573 aphasic trial participants being recorded as male. The mean age of participants in Big CACTUS was 65.4 years. Again, this is similar to the median age of 63 years for 5871 aphasic participants in the RELEASE data set.⁷⁷ This is also consistent with the average age seen in stroke rehabilitation trials (64.3 years) and is almost a decade younger than those seen by physicians in daily practice.⁷⁸

Computer use

The mean amount of self-managed practice with the computer therapy in Big CACTUS was 28 hours over 6 months. This is consistent with the amount of self-managed practice conducted in the pilot study: a mean of 25 hours over 5 months.²¹ In a recent systematic review of computerised aphasia therapy interventions, the protocols varied with recommended practice schedules of 10 to 11 hours over 2 months, 20 hours over 20 sessions, 24 hours over 11 to 12 weeks and 78 hours over 6 months.¹⁷ These studies were all conducted in clinical facilities with at least some therapist supervision, unlike Big CACTUS, which was a pragmatic study of computer therapy self-managed by the person with aphasia at home, encompassing greater participant control and choice over amounts of practice. The total amount of practice carried out by the Big CACTUS participants was similar to that in two of the studies described in the systematic review but over a longer duration of time. In addition, 61% of the Big CACTUS participants continued to practise beyond the 6-month supported self-managed trial period, suggesting that participants may choose to practise less intensively for a longer duration than provided in therapist-supervised protocols studied. In the Oral Reading for Language in Aphasia (ORLA) study,¹⁰ a study of computer-based reading therapy for people with aphasia, there was also some variation in the amount of practice participants chose to carry out.

Word-finding

The Big CACTUS trial demonstrated significant improvements in word-finding at the impairment level of 16.2% (95% CI 12.7% to 19.6%; $p < 0.0001$) in favour of CSLT compared with UC in the chronic phase (> 4 months) post stroke. This finding is very similar to the findings of the pilot study, based on 34 participants, which showed improvements of 19.8% (95% CI 4.4% to 35.2%; $p = 0.014$) in favour of the computer therapy compared with usual stimulation (support groups but no therapy intervention) more than 6 months post stroke. The Cochrane review⁴ included three studies that compared naming therapy to social support using impairment-based naming outcome measures. On pooling the data, no differences were seen (standardised mean difference 0.14, 95% CI -0.10 to 0.38; $p = 0.26$).⁴ However, the word-finding measures used were standardised and therefore will have been observing improvement to words that were not specifically treated in therapy. A small case series study (16 participants) by Best *et al.*⁵¹ suggested that only one in four participants improved on untreated words, and then they

improved by only a small amount (4%) between pre and post therapy measures. With control groups, the Big CACTUS trial showed no effect of the CSLT word-finding therapy on finding words that were not treated, in keeping with the findings in the 2012 Cochrane review of aphasia therapy⁷⁹ and previous studies of word-finding therapy.⁸⁰⁻⁸⁴

Functional communication, participation and quality of life

The CSLT for word-finding did not result in improvements at the level of communication activity and participation or quality of life in the Big CACTUS trial. A common criticism of impairment-based aphasia therapies is that improvements are not reflected in real-life, day-to-day communication.⁸⁵ The lack of generalisation of new word-finding ability to functional contexts is therefore consistent with findings from other, smaller studies of word-finding therapy. For example, Best *et al.*⁸⁵ demonstrated carryover of naming therapy into conversation for some individuals (as seen in Big CACTUS) but not for the group of study participants as a whole. This underlines the need for additional therapy components to be added to impairment-based intervention to aid use of new language skills.

Improvements in communication activity, participation or quality of life were not measured quantitatively in the CACTUS pilot study; however, some participants described functional improvements during qualitative interviews (e.g. use of their new words in conversation or in functional contexts, and improvements in confidence to communicate in functional communication settings).²² The quantitative findings from the full trial therefore diverge from these qualitative findings. One explanation of the differences could be that 'response shift' may occur in patient-rated outcomes, whereby the internal standards against which the participants rate themselves do not stay constant between one time point and another.⁸⁶ This could lead to descriptive improvements in interviews that are not accounted for in responses on patient-rated outcome measures.

Health economics

For the health economic evaluation, the base-case (primary) model-based cost-effectiveness analysis resulted in an ICER of £42,686 per QALY gained for CSLT compared with UC. This is much higher than the ICER estimated from the pilot study (£3127 per QALY).¹⁸ Reasons for this difference are that incremental costs associated with the intervention were lower in the pilot than in the main study (£469 compared with £733) and incremental QALYs were higher (0.15 compared with 0.02). In the pilot study, the intervention was delivered predominantly by a SLT who was part of the research team and more familiar with the computer program than the therapists delivering the intervention in the full, pragmatic trial in routine clinical practice conditions. In addition, only 48 bespoke words were prepared by the SLT in the pilot, whereas 100 words were prepared in the full trial. The amount of therapist time taken to deliver the intervention was therefore lower in the pilot (5 hours and 20 minutes rather than 9 hours) and had a lower cost. The full study has shown QALY measurement in this population with this intervention to be highly variable, which may account for the difference in QALY gains seen in the two studies. Of particular note is that the utility gain associated with a 'good response' was 0.07 (95% CI -0.15 to 0.29) in the pilot study, compared with 0.02 (95% CI -0.03 to 0.07) at 12 months in Big CACTUS. Caution must be taken with this comparison because the definition of a response differed between the two studies. However, clearly the CIs around the utility change have been reduced by Big CACTUS, but they have centred around the lower end of the interval estimated in the pilot study, which is also the area that overlaps zero. For this reason, the value of the uncertainty that remains is almost as high after Big CACTUS as it was after the pilot study: a population-level EVPI of £37.0M was estimated from the pilot study, compared with £34.6M based on Big CACTUS. This is due to the increased ICER and also the finding that we still cannot be certain about whether or not the intervention results in a QALY gain.

An additional factor is that the EQ-5D-5L was used in Big CACTUS, whereas the EQ-5D-3L was used in the pilot study. Research has shown that the EQ-5D-5L is likely to reduce the utility increment or decrement associated with a quality-of-life change, compared with the EQ-5D-3L, and in fact a NICE Decision Support Unit report on this topic⁸⁷ used the CACTUS pilot as a case study and estimated that

the utility gain associated with a good response would have been 0.02 had the EQ-5D-5L been used. However, in line with NICE recommendations,⁴⁸ in our base-case economic evaluation we used the van Hout *et al.*⁴⁹ cross-walk algorithm to map EQ-5D-5L responses onto the EQ-5D-3L tariff in order to generate utility scores; hence, the decrement associated with using the EQ-5D-5L should not be present. The utility gain associated with a good response was marginally lower when the EQ-5D-5L tariff was directly used, but the difference was marginal (0.017 compared with 0.02).

Discussion of using co-primary outcomes

Computerised speech and language therapy is predominantly impairment focused with recognition of the need to transfer any impairment-based gains to function with support, in this case conversation and functional activities with an SLTA or volunteer (it should be noted that adherence to the functional activities/conversational part of the intervention was low). It is commonplace in speech and language therapy to work on the impairment first so that patients can 'do' the task (in this case retrieving words). The new skills then need to generalise to function to be useful in everyday life. The majority of studies of aphasia have only impairment-based outcomes. In this trial, we recognised the importance of evaluating functional gains, but if we had only a functional outcome we would not know whether or not the impairment-focused intervention improved the impairment as intended and therefore would not know if it helped patients to retrieve words at any level. As both of these were considered clinically important to know, and the international aphasia community agrees that it is important to measure across the dimensions of the ICF, the trial was designed with co-primary outcomes. During trial design, we discussed what would be considered beneficial with both SLTs and people with aphasia. Impairment or functional gains or both were considered beneficial. Therefore, a Hochberg testing procedure prespecified that the intervention would be interpreted as beneficial if both outcomes were significant or if either outcome was very significant alone. The benefit of taking this approach is that it is possible to see whether or not the intervention shows benefit in any dimension of the ICF [i.e. just what the intervention primarily targets (impairment) or whether or not there is any desired carryover to functional communication (activity)]. The disadvantage of this approach is that, although interpretation is based on what was prespecified using the Hochberg procedure, the result in this case is not binary 'it works'/'it does not work', making it more complex to report. However, the finding that it is beneficial for improving word-finding impairment but does not generalise to function is useful. If only a functional outcome was used, the intervention would be rejected altogether, but this would be inappropriate given the marked improvements in impairment. Rather, the results of the co-primary outcomes indicate that the intervention needs to be built on to help impairment-based gains generalise to function.

Strengths of the trial

The Big CACTUS trial was a fully powered, pragmatic, multicentre RCT. The SLTs all worked in routine NHS clinical practice from where they recruited and treated the participants. The large number of sites (21 departments across 20 NHS trusts) included representation from all devolved nations of the UK, both urban and rural areas, and different NHS information technology services, which required a range of computer and software procurement strategies, policies and governance standards to be followed in the implementation of the CSLT intervention being tested in this trial. These real-world considerations support the generalisability of the trial findings.

A further strength of the trial is that the interventions were well described using the TIDieR template, and the CSLT intervention was manualised and is publicly available, enabling understanding and replication. Training on the intervention was also provided and described. The intervention was complex, including a skill mix of qualified SLTs to tailor the computer therapy to individual participant need and to train and monitor SLTAs or volunteers to provide lower-cost ongoing support. Fidelity, not only to treatment adherence but also to quality of intervention delivery, was measured to aid transparency of how well the intended complex intervention was delivered.

To our knowledge, Big CACTUS is the largest trial of computerised aphasia therapy to date. The trial recruited 278 participants to time and target, stopping seven participants short of the sample size calculation owing to a lower than predicted drop-out rate from the trial, requiring fewer participants to detect an effect of 10% on the primary outcome measures with the intended 90% power. The drop-out rate was only 9%, 6% lower than predicted from the pilot study. Fewer than one-third of trials meet their recruitment targets, with more than half requiring extensions.⁸⁸ We based our recruitment rate on that seen in the pilot study: one participant per month per site. We also built in contingency of having sufficient funding for four extra sites in addition to the 20 sites we predicted we needed should the recruitment rate be lower. The actual average recruitment rate was 0.8 participants per month per site (range 0.5–1.2). We also took into consideration that the participants all had aphasia and therefore it would take a longer time and more skill to ensure informed consent. This meant that the trial was planned based on realistic recruitment rates with a sufficient number of sites and recruitment time. The Big CACTUS trial therefore shows that trials with participants with post-stroke aphasia can be successfully conducted. The Big CACTUS team put considerable effort into recruitment and retention to the trial. Strategies included:

- funding 1 dedicated half-day per week of a SLT's time at each site to recruit participants
- a stepped approach to gaining informed consent from people with aphasia with differing needs of support, with different styles of information to suit different language abilities, all designed in collaboration with the trial PPI group
- use of a Consent Support Tool³¹ by SLTs with specialist skills in communication to indicate when to use which style of information
- including people identified as being unable to consent with support through carer/relative declarations
- monthly newsletters to encourage and maintain motivation for recruitment with a competitive element
- availability of a researcher centrally during office hours to answer questions and support recruiting therapists
- monthly aphasia-friendly newsletters thanking participants and informing them of the progress of the trial to help retention (these were also designed in collaboration with the PPI group).

Strengths of the trial design

The trial compared the CSLT intervention under study with UC, but also had a third, AC, group so any effect of CSLT seen over and above UC can be attributed to the speech and language therapy components of the intervention as opposed to the receipt of extra attention. There is debate about the use of appropriate AC in the aphasia literature, suggesting that often social support AC interventions are so closely matched to speech and language therapy (in terms of materials and support) that any difference between the interventions is eroded, making it more difficult to establish the effectiveness of one intervention over another.⁸⁹ In the CSLT intervention, owing to it taking a self-management approach, the extra attention gained by the participants is limited to the monthly support with computer exercises from volunteers or SLTAs. Therefore, our third group uniquely attempted to control for both additional activity (the focus of self-managed therapy) and additional attention. The attention provided was the same frequency as for the CSLT group, but was kept distinct from the attention received in the CSLT group by focusing on different materials (puzzle books).

Our eligibility criteria were inclusive and therefore representative of the population treated in practice. For example, it is standard practice to exclude people who have had more than one stroke in many stroke studies; however, eligibility was not limited to only having had one stroke in Big CACTUS to reflect that people who have had more than one stroke do routinely receive treatment. We included practical tests to see whether or not individuals were able to see and manipulate the computer software as part of our eligibility criteria. Similar screening is likely to be carried out in clinical practice. However, assumptions that older people struggle to use computers may reduce the number of older people screened using these practical tests in practice.

A range of outcome measures were used to explore the effect of CSLT on all dimensions of the ICF (impairment, activity and participation) and on quality of life. An international core outcome set for aphasia using these dimensions has been developed.^{24,90} Best *et al.*⁸⁵ identified that aphasia studies often measure only impairment-based improvement, stating 'attempts to measure carryover to everyday conversation are conspicuous by their absence'. This was supported by the outcomes of the 2016 Cochrane review⁴ for aphasia in which only 31 out of 71 randomised comparisons used a functional communication outcome for word-finding therapy. Although the co-primary outcome measures were conducted by qualified SLTs, secondary outcome measures considered the participant's own view and the views of carers using patient- and carer-rated outcome measures. An accessible variant of the EQ-5D-5L designed in collaboration with the PPI group enabled participants to record their own views regarding quality of life for the health economic analysis.

Random allocation to the trial groups was concealed and therefore not predictable for any trial staff. The conduct of blinded outcome measure assessments was a strength of the trial. (Our co-primary outcome measures were blinded, although the secondary patient-rated outcome measures could not be conducted blind to group allocation.) Blinding was relatively successful, with only 28 participants having at least one unblinded 6-month outcome measure conducted in total, which is negligible in terms of the numbers of assessments made in the trial. It is of note that the possibility of unblinding applied to only the co-primary word-finding measure as conversation videos were rated on the TOMs by SLTs independent of the trial, allowing blinding to group allocation and time point to be maintained. Reliability testing between SLTs scoring the conversations with the TOMs and those scoring the naming test at each site was a further strength of the design.

The trial benefited from comprehensive SAPs and health economic analysis plans being agreed before unblinding of the data for analysis (see www.journalslibrary.nihr.ac.uk/programmes/hta/122101/#). The trial was adequately designed and powered to address multiple primary and key research questions while controlling for the chances of making false-positive conclusions about the effects of the intervention. In addition, statistical analysis explored the impact of missing data on the results under a number of scenarios for sensitivity analysis, as well as the impact of adherence to the components of the CSLT and AC interventions.

The strengths of the health economic evaluation include its comprehensive investigation of many different scenarios and analyses. The use of an accessible version of the EQ-5D-5L is novel and permits the elicitation of utility scores from a patient group that is not well provided for with respect to preference-based utility measures.^{60,61} The collection of EQ-5D-5L questionnaires completed by carers by proxy and also for themselves allowed two additional valuable analyses to be conducted. Given the current uncertainty around how best to use the EQ-5D-5L measure, it was important to conduct analyses using all the valuation options. The economic evaluation undertaken alongside the CACTUS pilot study was simplistic and results were highly uncertain.¹⁸ Although this more complex evaluation has also produced highly uncertain results, we have gained important information on where the uncertainties lie. Finally, this evaluation has been conducted in line with best-practice guidelines and includes a rigorous investigation into the uncertainty associated with the decision-making problem.

Patient and public involvement

Patient and public involvement was a huge strength of this trial, meeting all four of its aims to:

1. facilitate the recruitment and inclusion of people with aphasia in the trial.
2. ensure that trial materials and processes were accessible to people with aphasia.
3. ensure that the interventions and trial procedures were appropriate and manageable for people with aphasia.
4. ensure that dissemination of trial results reached a broad audience in accessible formats.

Collaborating on the production of accessible information formats, including trial information, consent forms, wording to be used in trial procedures, computer-support guides, accessible outcome measures and monthly newsletters, has been described throughout the report. The PPI group led on the development of the study dissemination plan, considering all of the groups we need to target with the results, including people with aphasia, carers, SLTs, health professionals, guideline developers, commissioners, providers of services, voluntary sector organisations and researchers. The PPI group instigated and collaborated on the production of a film showing what aphasia is like, the CSLT intervention and our findings; a results booklet to accompany the film with key information facts and figures, and aphasia-friendly summaries of key points alongside; and flyers containing a simple summary of the key findings to send to all the participants of the trial and to display in public spaces in health and social care venues.

Limitations of the trial

The CSLT intervention has four key components: a computer with word-finding software on it, a SLT to set it up, independent practice by the person with aphasia and a volunteer/SLTA for support. There are a growing number of software options available that could be used as the 'software' component of this intervention and for consistency we limited this to just one option in the trial: StepByStep. Numerous apps are available at low cost; however, the majority of apps designed for word-finding have a limited range of words to practise, do not enable choice of personally relevant words or adding of personally relevant words and require the person with aphasia to make their own judgement about whether or not they said the correct word. They often also focus on naming at the word level and not use of words in sentences. Tactus Therapy (Tactus Therapy Solutions Ltd, Vancouver, BC, Canada) provides a suite of apps to help with word-finding, including 'Naming Therapy', which has a reasonable number of words available (418) and enables the addition of personally relevant words. The Tactus 'Advanced Naming Therapy' app focuses on use of words in sentences and conversations, which may help with generalisation of words to functional contexts. However, the Tactus apps do not provide feedback on whether or not the word has been named correctly. StepByStep is one of the most expensive word-finding software options available but has advantages over most other software in having thousands of words to choose from, the ability to add new words of personal importance and the provision of feedback on successful word-finding attempts. It also moves through a hierarchy of exercise difficulty to practise words in sentences.

Another limitation of working with software is that it evolves over time. StepByStep version 5 was new at the start of the trial and some instability in its functioning, particularly with the speech-recognition feedback function, increased the amount of time it took the SLTs to set it up, increased the time they needed to support participants with software problems and potentially limited exercises available to practice. These difficulties gradually resolved over the course of the trial. Only 33% of participants used their own computer, with 66% requiring a computer to be loaned by the NHS trust. This was because it was possible to install the software only on a PC, laptop or tablet running Microsoft Windows® 7 (Microsoft Corporation, Redmond, WA, USA) or above during the trial, restricting use of their own devices for users of Macs or iPads (Apple Inc., Cupertino, CA, USA). Since the trial ended, the software has become available for use on iPads, which will increase the number of people who could use their own device and automatically increases or decreases the level of exercise difficulty in response to user responses rather than relying on volunteers or assistants to identify the correct level of difficulty. Although the concept of the intervention remains unchanged, software may come down in price over time, exercises could be added to aid transfer to conversation and web-based options may open it up to a greater number of people using their own devices, etc. Such evolution may lead to changes in the cost and effectiveness of the intervention, meaning that, although the trial findings will remain informative, consideration will need to be given to the Big CACTUS results in relation to the time period in which the trial was conducted.

The AC group was intended to control for attention and activity carried out in the CSLT group. This was a difficult intervention to design and measure fidelity for. It is possible that puzzle books may not control for activity on a computer as well as if the puzzles were presented and completed on a computer. It may have also been easier to measure adherence to puzzles completed on the computer for a better comparison with the CSLT group adherence. The telephone calls were designed to control for attention received by volunteers/SLTAs in the CSLT group. Limitations of the telephone calls include that support was not face to face and was often received by the carer rather than directly by the person with aphasia.

Although the eligibility criteria aimed to be inclusive to represent treated populations in routine clinical practice, those who required treatment in a language other than English had to be excluded because the computer software was available only in English. In routine practice, guidelines urge SLTs to provide therapy in the language required by the patient.⁹¹ The inclusion criteria were kept broad as we have little knowledge of who benefits from CSLT. This can also be considered a limitation in some respects as the intervention was not targeted at those likely to be motivated to self-manage or likely to have good outcomes. Similarly, the trial design expected SLTs to set up therapy for 100 personally chosen words after randomisation to the CSLT group. In practice, it may be more likely that a SLT would see how motivated a patient is when using the software and whether or not any gains are indicated over a short trial period before investing the time in detailed set-up for a long duration.

Randomisation was stratified by severity of word-finding and by site, which was important to ensure that all sites had some participants in each of the three groups; this was particularly important so that UC was represented by all 20 NHS trusts. However, despite blocking of the allocation sequence, slightly more participants were randomised to the CSLT and UC groups than to the AC group. This was solely due to chance and not an error or subversion of the randomisation system. Too many sites recruiting small numbers of participants each and termination of recruitment after 278 patients may have contributed to this chance imbalance. The slight imbalance resulted in higher power than expected for the CSLT and UC primary comparison and slightly lower power than expected for the supportive CSLT and AC comparison, although power was adequate to address the intended research objectives.

The observed attrition, especially at 12 months, was slightly higher than anticipated, which may have limited our inference of the long-term effects of the intervention. However, we conducted sensitivity analyses to explore the impact of missing data at different time points.

As mentioned previously, 'attempts to measure carryover to everyday conversation are conspicuous by their absence'.⁸⁵ This is perhaps because it is difficult to know how best to capture changes and what changes to look for in conversation. We used the activity scale of the TOMs to rate the conversations elicited and recorded as it has good reliability and is likely to detect clinically meaningful change. However, concerns about the sensitivity of the measure have been raised as it is unlikely to detect small changes.⁹²

As patient participants were randomised, and any carers of these participants were invited to join the trial if they wished, carers were not randomised to the trial but were self-selecting. Outcomes for carers of people with aphasia were measured in the trial. Although characteristics were recorded for the patient participants, characteristics were not recorded for their carers.

In order to limit burden on the participants, it was not possible to measure all of the potential confounding variables (e.g. cognitive function and location of stroke using magnetic resonance imaging). Although randomisation was used to control for measured and unmeasured confounders, what we did not measure limited us from describing the trial population as fully as we would have liked, and hence from exploring potential heterogeneity of treatment effects in specific subgroups of potential unmeasured confounders.

A key limitation of the health economic evaluation is that base-case analyses rely on utility scores derived from an accessible version of the EQ-5D-5L that has not yet been validated beyond the face validity it has from being developed with people with aphasia. The need for a validated accessible version of the EQ-5D-5L for people with aphasia has been discussed in the literature.^{60,61} As there was no testing of how well people with aphasia understand the questions in the accessible EQ-5D-5L, it is possible that lack of understanding, particularly by participants with more severe aphasia, may have influenced the results. It appears that our use of the accessible EQ-5D-5L in our base-case analysis may have been conservative, given that cost-effectiveness results using carer proxy utility scores using a standard EQ-5D-5L questionnaire are more favourable for CSLT.

A further limitation of the health economic analysis is that only direct intervention costs are included. It is therefore assumed that there are no indirect resource use implications associated with CSLT, AC or UC. This was based on findings from the CACTUS pilot study, which found no important differences in indirect resource use between CSLT and UC, leading to a decision not to collect such data for the health economic analysis in Big CACTUS.^{18,93} However, some information on UC (in the form of SLT contact) was collected [see *Chapter 4, Usual-care speech and language therapy offered (fidelity/adherence to provision of usual care)*]. Mean SLT contact time reduced in the CSLT, UC and AC groups throughout the trial, and there may be an indication towards this reducing by slightly more (by approximately 0.5 to 1 hour) in the CSLT group than in the UC and AC groups. If this were the case, and there was an approximate 1 hour of SLT time cost saving associated with CSLT compared with UC and AC, the impact on the model-based base-case ICER would be relatively minor; an approximate £50 cost reduction for CSLT would reduce the ICER for CSLT compared with UC to approximately £40,000 per QALY gained.

Implications of the Big CACTUS trial findings

The implication of this trial is that people with aphasia can increase the number of hours of repetitive practice to improve word-finding by self-managing their practice of exercises tailored to their needs by a SLT and supported by an assistant or volunteer. This additional practice comes at a lower cost than if it was provided through an increase in face-to-face speech and language therapy.

The number of hours of repetitive practice achieved independently by people with aphasia leads to significant improvements in the ability to find words of personal importance and these improvements are maintained. Improvements in ability to find words of personal relevance were seen any time post stroke; therefore, time post stroke is not a barrier to learning new words with therapy.

However, the improvement in word-finding in the chronic phase (> 4 months) post stroke is limited to the words used in therapy and does not generalise to other words. It is therefore important that words used in therapy are chosen carefully to be personally relevant and therefore functionally useful for the lives of each individual being treated.

The aim of speech and language therapy is to improve the ability of people with aphasia to communicate in everyday situations and thus improve their participation in daily life with consequent increases in their quality of life. Although the Big CACTUS trial demonstrated significant improvements in word-finding, these improvements were seen only at the level of the impairment and did not lead to improvements in conversation or using the new words when given the opportunity in a functional context. It is therefore unsurprising that participants did not perceive improvements in their communication, participation or quality of life. The implication of this finding is that generalisation of impairment-based improvements may not occur with speech and language therapy without additional support.

The cost-effectiveness of CSLT remains uncertain; however, given the cost-effectiveness thresholds used by NICE in the UK, it is unlikely to be cost-effective for the whole group of people with aphasia. Subgroup analyses are prone to greater uncertainty than analyses of full trial populations, but our

analyses suggest that CSLT is more likely to be cost-effective for people with mild and moderate word-finding difficulties owing to a greater change in quality of life in these groups. For people with severe word-finding difficulties, our model-based analysis estimated quality-of-life improvements that were lower with computer therapy than with UC alone.

Further research

Areas of further research in order of priority include the following:

1. Investigating ways to assist with generalisation of newly learned vocabulary into use in conversation and other functional communication contexts. This should be informed by new knowledge of the neuroscience of recovery. The results of the Big CACTUS trial showed that some participants did use a few of their new words in conversation. It would be useful to explore the characteristics, personal and neurological, and behaviours of these people to compare with those who showed no generalisation to gain insights into what may help with generalisation.
2. Identification of what was practised and whether or not all of the exercises set up by the SLT were used. Currently, we have only looked at how much practice the participants of the study conducted. Further fidelity research into what was practised and whether or not all of the exercises set up by the SLT were used will help to further our understanding of the content of therapy practice that led to the Big CACTUS results and any changes that could usefully be made to the intervention.
3. Exploration of further cost and time efficiencies. The approach used in the Big CACTUS trial led to low-cost self-managed repetitive practice and consequent successful impairment-based word-finding improvement. Ways of making further cost and time efficiencies to explore include use of telehealth to set up and monitor exercises to save travel time, use of assistants/volunteers personalising the set of words to use in therapy instead of qualified SLTs and identification of people with aphasia who are motivated to use the software before investing time in setting it up.
4. Exploration of whom to target the intervention towards. We offered the intervention to a wide range of people with aphasia. It is important to investigate the characteristics of those who do well with the intervention and those who do not, in terms of demographic characteristics, personality and motivation, and the pattern of brain lesions resulting from the stroke.
5. Implementation of an optimised CSLT approach as part of NHS speech and language therapy provision. As a pragmatic trial, the Big CACTUS trial can offer insights into factors affecting implementation of self-managed computer therapy approaches within the NHS. Further research into the implementation of this approach as part of NHS speech and language therapy needs to be conducted to assist with making the approach available to people with aphasia for repetitive language practice components of their therapy provision.
6. Validation of the accessible variant of the EQ-5D-5L. Measurement of quality of life in people with aphasia for use in health economic analysis presents a challenge. We developed an accessible variant of the EQ-5D-5L for the Big CACTUS trial. We plan to carry out further validation of this tool. Development of accessible forms of other tools to measure quality-of-life changes resulting from communication therapies also needs to be carried out. In addition, finding out more about the utility benefit of the CSLT intervention would be highly valuable as the utility associated with achieving a good response to treatment was the most important parameter within the economic model.

Acknowledgements

Co-applicants: Dr Rebecca Palmer (Chief Investigator, University of Sheffield), Professor Cindy Cooper (Director, CTRU, University of Sheffield), Dr Nick Latimer (Senior Research Fellow in Health Economics, University of Sheffield), Professor Steven Julious (Senior Statistician, University of Sheffield), Professor Pam Enderby (Professor of Community Rehabilitation, University of Sheffield), Professor Audrey Bowen (Stroke Association John Marshall Memorial Professor, University of Manchester) and Professor Marian Brady (Professor of Stroke Care and Rehabilitation, Glasgow Caledonian University).

The Big CACTUS research team: Rebecca Palmer (Chief Investigator), Liz Cross (Trial Manager), Madeleine Harrison (Trial Researcher), Ellen Bradley (Trial Researcher), Helen Hughes, Helen Witts (Research SLTs), Tim Chater (Data and Information Systems Manager), Chris Turtle (Data Management Officer), Sarah Gonzalez, Katie Walker, Lauren O'Hara (Trial Support Officers), Kathryn MacKellar, Richard Simmonds (Administrative Assistants), Munyaradzi Dimairo (Trial Statistician), Nicholas Latimer (Senior Health Economist), Abualbishr Alshreef and Arjun Bhadhuri (Health Economists).

Trial Management Group: Dr Rebecca Palmer (Chief Investigator, University of Sheffield), Professor Cindy Cooper (Director, CTRU, University of Sheffield), Dr Nicholas Latimer (Senior Research Fellow in Health Economics, University of Sheffield), Professor Steven Julious (Senior Trial Statistician, University of Sheffield), Professor Pam Enderby (Emeritus Professor of Community Rehabilitation, University of Sheffield), Professor Audrey Bowen (Stroke Association John Marshall Memorial Professor, University of Manchester), Professor Marian Brady (Professor of Stroke Care and Rehabilitation, Glasgow Caledonian University), Dr Munyaradzi Dimairo (Trial Statistician, University of Sheffield), Elizabeth Cross (Trial Manager, University of Sheffield), Madeleine Harrison (Trial Researcher, University of Sheffield), Ellen Bradley (Trial Researcher, University of Sheffield), Helen Hughes (SLT, Sheffield Teaching Hospitals NHS Foundation Trust on secondment to University of Sheffield), Helen Witts (SLT, Derbyshire Community Health Services NHS Foundation Trust on secondment to University of Sheffield), Tim Chater (Data and Information Systems Manager, University of Sheffield), Mr Abualbishr Alshreef (Health Economist, University of Sheffield), Dr Arjun Bhadhuri (Health Economist, University of Sheffield) and Sarah Gonzalez (Trials Support Officer, University of Sheffield).

We gratefully acknowledge the hard work of the following: Jane and Peter Mortley, Steps Consulting Ltd, Stephanie Pinnell (administrative support, Norfolk Community Health and Care NHS Trust), Hazel Eaton (research support, Dorset Healthcare University NHS Foundation Trust), Maria Stevenson (administrative support, Livewell Southwest, Plymouth) and Leeds West National Research Ethics Committee.

We would like to thank the members of our PPI group for their invaluable advice and support: Gordon Burnett, Pamela Burnett, Ian Merriman, Traudel Merriman, Kate Sudworth, Tim Sudworth and Chris Welburn.

We would like to thank the following for their support of the trial: Paul Biagioni (Senior Manager, NI CRN), Martin Drummond (information Technology Support Services, Belfast Trust) and Rena Truscott (Livewell Southwest, Plymouth).

We offer special thanks to the members of our oversight committees, the Trial Steering Committee and Data Monitoring and Ethics Committee (DMEC): Professor Pip Logan (Chairperson and DMEC member, Professor of Rehabilitation, Queen's Medical Centre, Nottingham), Professor Peter Langhorne (Independent Stroke Specialist and DMEC member, Professor of Stroke Care, Glasgow Royal Infirmary), Dr Derrick Bennett (Independent Statistician and DMEC member, Senior Statistician, Clinical Trial Service Unit, University of Oxford), Mrs Elaine Roberts (Independent Stroke Specialist, The Stroke

ACKNOWLEDGEMENTS

Association, Director of Life After Stroke Services, Manchester), Ms Gail Paterson (Independent Speech Therapist, Specialist SLT, Pennine Care NHS Foundation Trust), Simon and Lesley Benneworth (Service User and Carer), Alan Hewitt (Service User), Professor Steven Julious (Trial Statistician, University of Sheffield), Rebecca Palmer (Chief Investigator, University of Sheffield), Elizabeth Cross (Trial Manager, University of Sheffield) and Madeleine Harrison (Trial Researcher, University of Sheffield).

Principal investigators: Ann-marie Anderson (NHS Ayrshire and Arran), Deborah Croall and Lucy Morgan (Cwm Taf Morgannwg University Health Board), Wendy Davies (NHS Greater Glasgow and Clyde), Niki Freedman and Mary Pointer (Essex Partnership University NHS Foundation Trust), Kath Gutteridge (Northamptonshire Healthcare NHS Foundation Trust), Amanda Harris and Helen Thorn (Sub-investigators) (Livewell Southwest, Plymouth), Lindsey Howat (Dorset Healthcare University NHS Foundation Trust), Heather Hurren (Somerset Partnership NHS Foundation Trust), Sally Knapp (Nottinghamshire Healthcare NHS Foundation Trust), Kay Martin (Cambridgeshire and Peterborough NHS Foundation Trust), Carolee McLaughlin and Crea Neeson (Belfast Health and Social Care Trust), Louise O'Neill and Cathy Magee (Northern Health and Social Care Trust), Anna Ray (Humber NHS Foundation Trust), Emma Rees and Sam Lloyd (Sub-investigators) (Abertawe Bro Morgannwg University Health Board, Swansea), Catherine Stamp (Northern Lincolnshire and Goole NHS Foundation Trust), Sonja Turner (Newcastle Upon Tyne Hospitals NHS Foundation Trust), Natascha Ullrich (Sheffield Teaching Hospitals NHS Foundation Trust), Heather Waldron and Nikki Milburn (City Hospitals Sunderland NHS Foundation Trust), Dorothy Wilson and Tammy Davidson-Thompson (Norfolk Community Health and Care NHS Trust) and Helen Witts (Derbyshire Community Health Services NHS Foundation Trust).

Outcome measure therapists: Nia Jenkins, Sam Lloyd, Sacha Price (Abertawe Bro Morgannwg University Health Board, Swansea), Fiona Burnett, Jenny Taylor (NHS Ayrshire and Arran), Linzi Pearson, Carolee McLaughlin (Belfast Health and Social Care Trust), Lissa Bage, Lisa Cracknell, Gill Housley (Cambridgeshire and Peterborough NHS Foundation Trust), Sarah Barker, Rachel Barry, Charlotte Craig, Emily Osborne (City Hospitals Sunderland NHS Foundation Trust), Jane Canning, Laura Waring (Cwm Taf Morgannwg University Health Board), Rachel Bull, Barbara Cox, Fiona Mann (Derbyshire Community Health Services NHS Foundation Trust), Lauren Guy, Molly Sugrue, Mary Wells-West (Dorset Healthcare University NHS Foundation Trust), Monique Hinds, Niamh Kelly, Pippa Johnson (Essex Partnership University NHS Foundation Trust), Audrey Williamson, Angela Moar (NHS Greater Glasgow and Clyde), Emma Macleod, Belinda Smith (Humber NHS Foundation Trust), Alice Bowery, Claire Gatehouse, Philippa Knox (Livewell Southwest, Plymouth), Rosie Robertson, Jill Summersall (Newcastle Upon Tyne Hospitals NHS Foundation Trust), Miriam Horne, Anna Money, Janine Short (Norfolk Community Health and Care NHS Trust), Gill Hood, Catherine Macpherson (Northamptonshire Healthcare NHS Foundation Trust), Jennifer Handforth, Cathy Magee, Nicola McCarron, Stephanie Murray (Northern Health and Social Care Trust), Nicola Crook, Gemma Green, Emily Lomas, Elizabeth Marshall, Joanna Taylor (Northern Lincolnshire and Goole NHS Foundation Trust), Becky Thurman, Kirsten Hilton, Barbara Wilkinson (Nottinghamshire Healthcare NHS Foundation Trust), Sarah Minshall, Sarah Ross (Sheffield Teaching Hospitals NHS Foundation Trust), Sarah Dunne, Jessica Hawkes and Rachel Lee (Somerset Partnership NHS Foundation Trust).

Additional funding

We would like to thank The Tavistock Trust for Aphasia for the additional funding contribution that allowed us to, after the trial, provide the opportunity for participants not randomised to the computer intervention to experience the use of the software.

Contributions of authors

Rebecca Palmer (<https://orcid.org/0000-0002-2335-7104>) (Reader in Communication and Stroke Rehabilitation) was involved in conceiving and designing the work, the acquisition of data, analysis and management of data, interpretation of data and producing the report.

Munyaradzi Dimairo (<https://orcid.org/0000-0002-9311-6920>) (Research Fellow, Medical Statistics) was involved in conceiving and designing the work, analysis and management of data and producing the report.

Nicholas Latimer (<https://orcid.org/0000-0001-5304-5585>) (Reader in Health Economics) was involved in conceiving and designing the work, analysis and management of data, interpretation of data and producing the report.

Elizabeth Cross (<https://orcid.org/0000-0002-7976-8463>) (Research Fellow) was involved in acquisition of data and producing the report.

Marian Brady (<https://orcid.org/0000-0002-4589-7021>) (Professor of Stroke Care and Rehabilitation) was involved in conceiving and designing the work, interpretation of data and producing the report.

Pam Enderby (<https://orcid.org/0000-0002-4371-9053>) (Emeritus Professor of Community Rehabilitation) was involved in conceiving and designing the work, interpretation of data and producing the report.

Audrey Bowen (<https://orcid.org/0000-0003-4075-1215>) (Stroke Association John Marshall Memorial Professor of Neuropsychological Rehabilitation) was involved in conceiving and designing the work, interpretation of data and producing the report.

Steven Julious (<https://orcid.org/0000-0002-9917-7636>) (Professor of Medical Statistics) was involved in conceiving and designing the work, analysis and management of data and interpretation of data.

Madeleine Harrison (<https://orcid.org/0000-0001-9874-921X>) (Research Associate) was involved in acquisition of data and producing the report.

Abualbishr Alshreef (<https://orcid.org/0000-0003-2737-1365>) (Research Fellow, Health Economics) was involved in analysis and management of data and producing the report.

Ellen Bradley (<https://orcid.org/0000-0002-4147-3627>) (Research Assistant) was involved in acquisition of data and producing the report.

Arjun Bhadhuri (<https://orcid.org/0000-0003-1220-0731>) (Research Associate, Health Economics) was involved in analysis and management of data and producing the report.

Tim Chater (<https://orcid.org/0000-0002-1138-0147>) (Data and Information Systems Manager) was involved in analysis and data management.

Helen Hughes (<https://orcid.org/0000-0002-1069-9922>) (SLT) was involved in the acquisition of data for the work.

Helen Witts (<https://orcid.org/0000-0001-5177-7117>) (SLT) was involved in the acquisition of data for the work.

ACKNOWLEDGEMENTS

Esther Herbert (<https://orcid.org/0000-0002-1224-5457>) (National Institute for Health Research Research Methods Fellow) was involved in analysis and data management.

Cindy Cooper (<https://orcid.org/0000-0002-2995-5447>) (Professor of Health Services Research and Clinical Trials) was involved in conceiving and designing the work, interpretation of data and producing the report.

Publications

Palmer R, Cooper C, Enderby P, Brady M, Julious S, Bowen A, *et al.* Clinical and cost effectiveness of computer treatment for aphasia post stroke (Big CACTUS): study protocol for a randomised controlled trial. *Trials* 2015;**16**:18.

Palmer R, Hughes H, Chater T. What do people with aphasia want to be able to say? A content analysis of words identified as personally relevant by people with aphasia. *PLoS One* 2017;**12**:e0174065.

Palmer R, Witts H, Chater T. What speech and language therapy do community dwelling stroke survivors with aphasia receive in the UK? *PLoS One* 2018;**13**:e0200096.

Palmer R, Dimairo M, Cooper C, Enderby P, Brady M, Bowen A, *et al.* 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. *Lancet Neurol* 2019;**18**:821–33.

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Please note that exclusive use will be retained until the publication of major outputs. Access to anonymised data may be granted following review.

Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: <https://understandingpatientdata.org.uk/data-citation>.

References

1. Laska AC, Hellblom A, Murray V, Kahan T, Von Arbin M. Aphasia in acute stroke and relation to outcome. *J Intern Med* 2001;**249**:413–22. <https://doi.org/10.1046/j.1365-2796.2001.00812.x>
2. Kauhanen ML, Korpelainen JT, Hiltunen P, Määttä R, Mononen H, Brusin E, *et al.* Aphasia, depression, and non-verbal cognitive impairment in ischaemic stroke. *Cerebrovasc Dis* 2000;**10**:455–61. <https://doi.org/10.1159/000016107>
3. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. *J Speech Lang Hear Res* 2008;**51**:S225–39. [https://doi.org/10.1044/1092-4388\(2008/018\)](https://doi.org/10.1044/1092-4388(2008/018))
4. Brady MC, Kelly H, Godwin J, Enderby P, Campbell P. Speech and language therapy for aphasia following stroke. *Cochrane Database Syst Rev* 2016;**6**:CD000425. <https://doi.org/10.1002/14651858.CD000425.pub4>
5. Allen L, Mehta S, McClure JA, Teasell R. Therapeutic interventions for aphasia initiated more than six months post stroke: a review of the evidence. *Top Stroke Rehabil* 2012;**19**:523–35. <https://doi.org/10.1310/tsr1906-523>
6. Breitenstein C, Grewe T, Flöel A, Ziegler W, Springer L, Martus P, *et al.* 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. *Lancet* 2017;**389**:1528–38. [https://doi.org/10.1016/S0140-6736\(17\)30067-3](https://doi.org/10.1016/S0140-6736(17)30067-3)
7. Kennedy N. *Results of NI RCSLT Survey of Communication Needs After Stroke*. London: Royal College of Speech and Language Therapists; 2018. URL: www.rcslt.org/governments/ni_docs/nircsltsurveycommneeds2018 (accessed 18 June 2018).
8. Bowen A, Hesketh A, Patchick E, Young A, Davies L, Vail A, *et al.* Clinical effectiveness, cost-effectiveness and service users' perceptions of early, well-resourced communication therapy following a stroke: a randomised controlled trial (the ACT NoW Study). *Health Technol Assess* 2012;**16**(26). <https://doi.org/10.3310/hta16260>
9. David R, Enderby P, Bainton D. Treatment of acquired aphasia: speech therapists and volunteers compared. *J Neurol Neurosurg Psychiatry* 1982;**45**:957–61. <https://doi.org/10.1136/jnnp.45.11.957>
10. Cherney LR. Oral Reading for Language in Aphasia (ORLA): evaluating the efficacy of computer-delivered therapy in chronic nonfluent aphasia. *Top Stroke Rehabil* 2010;**17**:423–31. <https://doi.org/10.1310/tsr1706-423>
11. Corwin M, Wells M, Koul R, Dembowski J. Computer-assisted anomia treatment for persons with chronic aphasia: generalization to untrained words. *J Med Speech Lang Pathol* 2014;**21**:149–63.
12. Kurland J, Baldwin K, Tauer C. Treatment-induced neuroplasticity following intensive naming therapy in a case of chronic Wernicke's aphasia. *Aphasiology* 2010;**24**:737–51. <https://doi.org/10.1080/02687030903524711>
13. Fink RB, Brecher A, Schwartz MF, Robey RR. A computer-implemented protocol for treatment of naming disorders: evaluation of clinician-guided and partially self-guided instruction. *Aphasiology* 2002;**16**:1061–86. <https://doi.org/10.1080/02687030244000400>
14. Mortley J, Wade J, Enderby P. Superhighway to promoting a client-therapist partnership? Using the internet to deliver word-retrieval computer therapy, monitored remotely with minimal speech and language therapy input. *Aphasiology* 2004;**18**:193–211. <https://doi.org/10.1080/02687030344000553>

15. Van de Sandt-Koenderman WME. Aphasia rehabilitation and the role of computer technology: can we keep up with modern times? *Int J Speech Lang Pathol* 2011;**13**:21–7. <https://doi.org/10.3109/17549507.2010.502973>
16. Stark BC, Warburton EA. Improved language in chronic aphasia after self-delivered iPad speech therapy. *Neuropsychol Rehabil* 2018;**28**:818–31. <https://doi.org/10.1080/09602011.2016.1146150>
17. Zheng C, Lynch L, Taylor N. Effect of computer therapy in aphasia: a systematic review. *Aphasiology* 2016;**30**:211–44.
18. Latimer NR, Dixon S, Palmer R. Cost-utility of self-managed computer therapy for people with aphasia. *Int J Technol Assess Health Care* 2013;**29**:402–9. <https://doi.org/10.1017/S0266462313000421>
19. Palmer R, Mortley J. How I offer impairment therapy (2): from idealism to realism, step by step. *Speech & Language Therapy in Practice* 2011;**Winter**:29–32.
20. Steps Consultancy Ltd. *StepByStep*© Software. URL: www.aphasia-software.com (accessed 18 June 2018).
21. Palmer R, Enderby P, Cooper C, Latimer N, Julious S, Paterson G, *et al*. Computer therapy compared with usual care for people with long-standing aphasia poststroke: a pilot randomized controlled trial. *Stroke* 2012;**43**:1904–11. <https://doi.org/10.1161/STROKEAHA.112.650671>
22. Palmer R, Enderby P, Paterson G. Using computers to enable self-management of aphasia therapy exercises for word finding: the patient and carer perspective. *Int J Lang Commun Disord* 2013;**48**:508–21. <https://doi.org/10.1111/1460-6984.12024>
23. World Health Organization (WHO). *International Classification of Functioning, Disability and Health*. Geneva: WHO; 2001. URL: www.who.int/classifications/icf/en/ (accessed 15 February 2019).
24. Wallace SJ, Worrall L, Rose T, Le Dorze G, Breitenstein C, Hilari K, *et al*. A core outcome set for aphasia treatment research: the ROMA consensus statement. *Int J Stroke* 2019;**14**:180–5. <https://doi.org/10.1177/1747493018806200>
25. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, *et al*. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;**340**:c869. <https://doi.org/10.1136/bmj.c869>
26. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Ann Intern Med* 2017;**167**:40–7. <https://doi.org/10.7326/M17-0046>
27. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, *et al*. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;**337**:a2390. <https://doi.org/10.1136/bmj.a2390>
28. Ioannidis JP, Evans SJ, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D, CONSORT Group. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;**141**:781–8. <https://doi.org/10.7326/0003-4819-141-10-200411160-00009>
29. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, *et al*. Better reporting of interventions: Template for Intervention Description and Replication (TIDieR) checklist and guide. *BMJ* 2014;**348**:g1687. <https://doi.org/10.1136/bmj.g1687>
30. Swinburn K, Porter G, Howard D. *Comprehensive Aphasia Test*. Hove: Psychology Press; 2004. <https://doi.org/10.1037/t13733-000>
31. Jayes M, Palmer R. *Consent Support Tool: Including People with Communication Disorders in Health Research Studies*. Macclesfield: Napier Hill Press; 2016.

32. Snell C, Sage K, Lambon Ralph MA. How many words should we provide in anomia therapy? A meta-analysis and a case series study. *Aphasiology* 2010;**24**:1064–94. <https://doi.org/10.1080/02687030903372632>
33. Enderby P, John A, Petheram B. *Therapy Outcome Measures for Rehabilitation Professionals: Speech and Language Therapy, Physiotherapy, Occupational Therapy*. 2nd edn. Chichester: John Wiley & Sons; 2013.
34. Long A, Hesketh A, Paszek G, Booth M, Bowen A. Development of a reliable self-report outcome measure for pragmatic trials of communication therapy following stroke: the Communication Outcome after Stroke (COAST) scale. *Clin Rehabil* 2008;**22**:1083–94. <https://doi.org/10.1177/0269215508090091>
35. Julious S. Letter to the editors. *Biometrics* 2004;**60**:284. https://doi.org/10.1111/j.0006-341X.2004.171_1.x
36. Julious SA, Owen RJ. Sample size calculations for clinical studies allowing for uncertainty about the variance. *Pharm Stat* 2006;**5**:29–37. <https://doi.org/10.1002/pst.197>
37. Group ICoHEEW. Statistical principles for clinical trials: ICH harmonized tripartite guideline. *Stat Med* 1999;**18**:1905–42.
38. StataCorp. *Stata Statistical Software*. In: Special Edition Release. College Station, TX: StataCorp LLC; 2017.
39. van Buuren S, Oudshoorn CGM. *Multivariate Imputation by Chained Equations: MICE V1.0 User's Manual*. Leiden: TNO; 2000.
40. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007;**16**:219–42. <https://doi.org/10.1177/0962280206074463>
41. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;**30**:377–99. <https://doi.org/10.1002/sim.4067>
42. Colantuoni E, Scharfstein DO, Wang C, Hashem MD, Leroux A, Needham DM, Girard TD. Statistical methods to compare functional outcomes in randomized controlled trials with high mortality. *BMJ* 2018;**360**:j5748. <https://doi.org/10.1136/bmj.j5748>
43. Hochberg Y, Tamhane A. Designing Experiments for Multiple Comparisons. In Hochberg Y, Tamhane AC, editors. *Multiple Comparison Procedures*. Hoboken, NJ: John Wiley & Sons Inc; 1987. <https://doi.org/10.1002/9780470316672>
44. Bowen A, Hesketh A, Long A, Patchick E. *Scoring Instructions for COAST and CaCOAST*. 2009. URL: www.click2go.umip.com/i/coa/coast.html (accessed 8 August 2016).
45. Long A, Hesketh A, Bowen A, ACT NoW Research Study. Communication outcome after stroke: a new measure of the carer's perspective. *Clin Rehabil* 2009;**23**:846–56. <https://doi.org/10.1177/0269215509336055>
46. Van Reenen M, Janssen B. *EQ-5D-5L User Guide: Basic Information on how to use the EQ-5D-5L Instrument*. 2015. URL: https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf (accessed 15 February 2017).
47. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: a EQ-5D-5L value set for England. *Health Econ* 2018;**27**:7–22. <https://doi.org/10.1002/hec.3564>
48. National Institute for Health and Care Excellence (NICE). *Position Statement on use of the EQ-5D-5L Valuation Set*. London: NICE; 2017. URL: www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisal-guidance/eq5d5l_nice_position_statement.pdf (accessed 18 June 2018).

49. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, *et al.* Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;**15**:708–15. <https://doi.org/10.1016/j.jval.2012.02.008>
50. Palmer R, Dimairo M, Cooper C, Enderby P, Brady M, Bowen A, *et al.* 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. *Lancet Neurol* 2019;**18**:821–33. [https://doi.org/10.1016/S1474-4422\(19\)30192-9](https://doi.org/10.1016/S1474-4422(19)30192-9)
51. Best W, Greenwood A, Grassly J, Herbert R, Hickin J, Howard D. Aphasia rehabilitation: does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex* 2013;**49**:2345–57. <https://doi.org/10.1016/j.cortex.2013.01.005>
52. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal 2013. Process and Methods [PMG9]*. London: NICE; 2013. URL: www.nice.org.uk/article/pmg9/chapter/foreword (accessed 18 June 2018).
53. Briggs AH, Gray AM. Methods in health service research: handling uncertainty in economic evaluations of healthcare interventions. *BMJ* 1999;**319**:635. <https://doi.org/10.1136/bmj.319.7210.635>
54. Drummond M, Brandt A, Luce B, Rovira J. Standardizing methodologies for economic evaluation in health care. Practice, problems, and potential. *Int J Technol Assess Health Care* 1993;**9**:26–36. <https://doi.org/10.1017/s0266462300003007>
55. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ* 2005;**14**:487–96. <https://doi.org/10.1002/hec.944>
56. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ* 2004;**13**:461–75. <https://doi.org/10.1002/hec.843>
57. Brønnum-Hansen H, Davidsen M, Thorvaldsen P, Danish MONICA Study Group. Long-term survival and causes of death after stroke. *Stroke* 2001;**32**:2131–6. <https://doi.org/10.1161/hs0901.094253>
58. Office for National Statistics (ONS). *National Life Tables, UK: 2014 to 2016*. Newport: ONS; 2017. URL: www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2014to2016 (accessed 18 June 2018).
59. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;**13**:509–18. <https://doi.org/10.1111/j.1524-4733.2010.00700.x>
60. Whitehurst DG, Latimer NR, Kagan A, Palmer R, Simmons-Mackie N, Hoch JS. Preference-based health-related quality of life in the context of aphasia: a research synthesis. *Aphasiology* 2015;**29**:763–80. <https://doi.org/10.1080/02687038.2014.985581>
61. Whitehurst DGT, Latimer NR, Kagan A, Palmer R, Simmons-Mackie N, Victor JC, Hoch JS. Developing accessible, pictorial versions of health-related quality-of-life instruments suitable for economic evaluation: a report of preliminary studies conducted in Canada and the United Kingdom. *Pharmacoecon Open* 2018;**2**:225–31. <https://doi.org/10.1007/s41669-018-0083-2>
62. Al-Janabi H, van Exel J, Brouwer W, Coast J. A framework for including family health spillovers in economic evaluation. *Med Decis Making* 2016;**36**:176–86. <https://doi.org/10.1177/0272989X15605094>
63. Hernandez-Alava M, Wailoo A, Pudney S. *Methods for Mapping Between the EQ-5D-5L and the 3L for Technology Appraisal*. London: NICE Decision Support Unit; 2017. URL: <http://nicedsu.org.uk/wp-content/uploads/2017/05/Mapping-5L-to-3L-DSU-report.pdf> (accessed 18 June 2018).

64. Pickard AS, Johnson JA, Feeny DH, Shuaib A, Carriere KC, Nasser AM. Agreement between patient and proxy assessments of health-related quality of life after stroke using the EQ-5D and Health Utilities Index. *Stroke* 2004;**35**:607–12. <https://doi.org/10.1161/01.STR.0000110984.91157.BD>
65. Hernandez M, Pudney S. EQ5DMAP: a command for mapping between EQ-5D-3L and EQ-5D-5L. *Stata J* 2017;**18**:395–415. <https://doi.org/10.1177/1536867X1801800207>
66. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Med Res Methodol* 2014;**14**:75. <https://doi.org/10.1186/1471-2288-14-75>
67. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: John Wiley & Sons; 2004.
68. Curtis LA, Burns A. *Unit Costs of Health and Social Care 2017*. Canterbury: PSSRU, University of Kent; 2017.
69. GOV.UK. *Expenses and Benefits: Business Travel Mileage for Employees' Own Vehicles*. 2018. URL: www.gov.uk/expenses-and-benefits-business-travel-mileage/rules-for-tax (accessed 18 June 2018).
70. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet* 2002;**360**:711–15. [https://doi.org/10.1016/S0140-6736\(02\)09832-X](https://doi.org/10.1016/S0140-6736(02)09832-X)
71. Griffin S, Welton NJ, Claxton K. Exploring the research decision space: the expected value of information for sequential research designs. *Med Decis Making* 2010;**30**:155–62. <https://doi.org/10.1177/0272989X09344746>
72. Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. *Med Decis Making* 2014;**34**:311–26. <https://doi.org/10.1177/0272989X13505910>
73. Zellner A, Huang DS. Further properties of efficient estimators for seemingly unrelated regression equations. *Int Econ Rev* 1962;**3**:300–13. <https://doi.org/10.2307/2525396>
74. Alshreef A, Wailoo AJ, Brown SR, Tiernan JP, Watson AJM, Biggs K, et al. Cost-Effectiveness of haemorrhoidal artery ligation versus rubber band ligation for the treatment of grade II-III haemorrhoids: analysis using evidence from the HubBLE trial. *Pharmacoecon Open* 2017;**1**:175–84. <https://doi.org/10.1007/s41669-017-0023-6>
75. Thomas KS, Bradshaw LE, Sach TH, Batchelor JM, Lawton S, Harrison EF, et al. Silk garments plus standard care compared with standard care for treating eczema in children: A randomised, controlled, observer-blind, pragmatic trial (CLOTHES Trial). *PLOS Med* 2017;**14**:e1002280. <https://doi.org/10.1371/journal.pmed.1002280>
76. Wolfe CD, Crichton SL, Heuschmann PU, McKeivitt CJ, Toschke AM, Grieve AP, Rudd AG. Estimates of outcomes up to ten years after stroke: analysis from the prospective South London Stroke Register. *PLOS Med* 2011;**8**:e1001033. <https://doi.org/10.1371/journal.pmed.1001033>
77. Ali M on behalf of the RELEASE collaboration. *Establishing an International Database of 5932 Individual Participant's Data to Inform the RELEASE Project*. Paper presented at the 4th European Stroke Organisation Conference, Gothenburg, Sweden, 2018.
78. Gaynor EJ, Geoghegan SE, O'Neill D. Ageism in stroke rehabilitation studies. *Age Ageing* 2014;**43**:429–31. <https://doi.org/10.1093/ageing/afu026>
79. Brady MC, Kelly H, Godwin J, Enderby P. Speech and language therapy for aphasia following stroke. *Cochrane Database Syst Rev* 2012;**5**:CD000425. <https://doi.org/10.1002/14651858.CD000425.pub3>
80. Nickels L. Therapy for naming disorders: revisiting, revising, and reviewing. *Aphasiology* 2002;**16**:935–79. <https://doi.org/10.1080/02687030244000563>

REFERENCES

81. Abel S, Schultz A, Radermacher I, Willmes K, Huber W. Decreasing and increasing cues in naming therapy for aphasia. *Aphasiology* 2005;**19**:831–48. <https://doi.org/10.1080/02687030500268902>
82. Fillingham JK, Sage K, Lambon Ralph MA. The treatment of anomia using errorless learning. *Neuropsychol Rehabil* 2006;**16**:129–54. <https://doi.org/10.1080/09602010443000254>
83. Laganaro M, Di Pietro M, Schnider A. Computerised treatment of anomia in chronic and acute aphasia: an exploratory study. *Aphasiology* 2003;**17**:709–21. <https://doi.org/10.1080/02687030344000193>
84. Wisenburn B, Mahoney K. A meta-analysis of word-finding treatments for aphasia. *Aphasiology* 2009;**23**:1338–52. <https://doi.org/10.1080/02687030902732745>
85. Best W, Grassly J, Greenwood A, Herbert R, Hickin J, Howard D. A controlled study of changes in conversation following aphasia therapy for anomia. *Disabil Rehabil* 2011;**33**:229–42. <https://doi.org/10.3109/09638288.2010.534230>
86. Schwartz CE, Andresen EM, Nosek MA, Krahn GL, RRTC Expert Panel on Health Status Measurement. Response shift theory: important implications for measuring quality of life in people with disability. *Arch Phys Med Rehabil* 2007;**88**:529–36. <https://doi.org/10.1016/j.apmr.2006.12.032>
87. Wailoo A, Hernandez-Alava M, Grimm S, Pudney S, Gomes M, Sadique Z, et al. *Comparing the EQ-5D-3L and 5L Versions. What are the Implications for Cost Effectiveness Estimates?* Sheffield: Decision Support Unit, School of Health and Related Research, University of Sheffield; 2017. URL: http://nicedsu.org.uk/wp-content/uploads/2017/05/DSU_3L-to-5L-FINAL.pdf (accessed 15 April 2020).
88. McGill K, Brady M, Godwin J, Sackley C. Efficiency of recruitment to stroke rehabilitation randomised controlled trials: secondary analysis of recruitment data. Paper presented at the International Clinical Trials Methodology Conference, Liverpool, UK, 7–10 May 2017, abstract no. 86.
89. Brady MC, Godwin J, Kelly H, Enderby P, Elders A, Campbell P. Attention control comparisons with SLT for people with aphasia following stroke: methodological concerns raised following a systematic review. *Clin Rehabil* 2018;**32**:1383–95. <https://doi.org/10.1177/0269215518780487>
90. Wallace SJ, Worrall L, Rose T, Le Dorze G. Using the International Classification of Functioning, Disability, and Health to identify outcome domains for a core outcome set for aphasia: a comparison of stakeholder perspectives. *Disabil Rehabil* 2019;**41**:564–73. <https://doi.org/10.1080/09638288.2017.1400593>
91. Royal College of Speech & Language Therapists. *RCSLT Guidance to Support Members to Adhere to the HCPC Standards*. London: Royal College of Speech & Language Therapists; 2016. URL: www.rcslt.org/cq_live/communication/rcslt_guidance/rcslt_guidance (accessed 26 June 2018).
92. John A. Therapy outcome measures: where are we now? *Int J Speech Lang Pathol* 2011;**13**:36–42. <https://doi.org/10.3109/17549507.2010.497562>
93. Palmer R, Cooper C, Enderby P, Brady M, Julious S, Bowen A, Latimer N. Clinical and cost effectiveness of computer treatment for aphasia post stroke (Big CACTUS): study protocol for a randomised controlled trial. *Trials* 2015;**16**:18. <https://doi.org/10.1186/s13063-014-0527-7>
94. Stroke Association. *Aphasia and its Effects*. Stroke Association. URL: www.stroke.org.uk/what-is-stroke/what-is-aphasia/aphasia-and-its-effects (accessed 26 April 2018).
95. Royal College of Speech & Language Therapists. *RCSLT Resource Manual for Commissioning and Planning Services for SLCN: Aphasia*. Royal College of Speech and Language Therapists; 2009. URL: www.rcslt.org/speech_and_language_therapy/commissioning/aphasia_plus_intro (accessed 26 April 2018).

Appendix 1 List of participating NHS trusts

- Sheffield Teaching Hospitals NHS Foundation Trust.
- Humber NHS Foundation Trust.
- Newcastle Upon Tyne Hospitals NHS Foundation Trust.
- Northern Health and Social Care Trust.
- Belfast Health and Social Care Trust.
- Essex Partnership University NHS Foundation Trust.
- NHS Greater Glasgow and Clyde.
- Cwm Taf Morgannwg University Health Board.
- Derbyshire Community Health Services NHS Trust.
- Nottinghamshire Healthcare NHS Trust.
- Northern Lincolnshire and Goole NHS Foundation Trust.
- Livewell Southwest (Plymouth Community Healthcare).
- Norfolk Community Health and Care NHS Trust.
- Somerset Partnership NHS Foundation Trust.
- Cambridgeshire and Peterborough NHS Foundation Trust.
- Northamptonshire Healthcare NHS Foundation Trust.
- Dorset HealthCare University NHS Foundation Trust.
- City Hospitals Sunderland NHS Foundation Trust.
- Abertawe Bro Morgannwg University Health Board.
- NHS Ayrshire and Arran.

Appendix 2 Process for selection of Therapy Outcome Measures raters and scoring procedure

The TOMs is one of the co-primary outcome measures for Big CACTUS. It is a rating of participants' ability to communicate in conversation. Each participant will have a 10-minute video-recorded conversation at baseline and 6, 9 and 12 months with a SLT (different therapists at different time points). Some videos recorded are longer than the advised 10 minutes. Two research SLTs rated a sample of videos based on the first 10 minutes and the whole video. The scores given after watching the whole video were the same as those given after watching only the first 10 minutes; therefore, only the first 10 minutes of each video will be used for rating purposes. The videos will be rated by independent SLTs using the activity scale of the aphasia TOM. As there will be > 1000 videos to rate, we need a number of raters.

Our primary comparison is the difference in change in scores between time points between groups. It is proposed that all of the videos at all time points of all participants at each site will be rated by the same rater. Therefore, it is important to know that raters have good intrarater reliability to give us confidence that any change in the scores across time points is likely to be due to change in the participant's communication ability and not due to variability in a rater's scoring (intrarater reliability). The raw scores will also be presented at each time point and data will be adjusted for baseline differences between groups. To ensure that the differences between scores are a true reflection of the differences in scores between participants, we need to minimise the variation in scoring between raters (inter-rater reliability).

The TOMs scale has six descriptors corresponding to 0, 1, 2, 3, 4 and 5. It is an 11-point scale as scores between two descriptors can be given (e.g. 1.5, 3.5). Previous work conducted with the TOMs suggests that the measure can be used reliably and that a variation of 0.5 in scores of different raters is acceptable.

To ensure confidence in the scoring of the raters who were used to rate the videos in the Big CACTUS trial, 18 raters wishing to participate attended a benchmarking session in which a sample of videos were watched and scores were discussed. Discrepancies in interpretation of the scoring system were highlighted and resolved. An agreed set of scoring instructions was compiled. The 18 raters agreed that 0.5 would be an acceptable amount of variation in the scores. The 18 raters then used the refined scoring instructions alongside the TOMs activity scale to rate 10 videos (five pairs of videos from five different participants at different time points). All 18 raters rated all of the same 10 videos (time point 1). After 6 weeks, the same 10 videos were sent back to the raters in a different order. Seventeen of the 18 original raters provided ratings of the videos (time point 2).

We need to understand the intrarater and inter-rater reliability of the rating given by these raters in order to have confidence in the scores provided by them, which constitute our co-primary outcome measure. Raters will be used to score videos for the study only if their intrarater reliability is $\geq 70\%$ (of scores within 0.5 between time points 1 and 2). We also need to ensure that raters are rating in a similar way to each other. As there is no gold standard rater, we have taken the median as the benchmark of 'what most people think the score should be'. We have included only raters that rate $\geq 70\%$ of their scores within 0.5 of the median, taking into consideration time points 1 and 2.

The intraclass correlation coefficients take account of the difference between scores at time points 1 and 2 (not only whether they are within 0.5 or not). Although all raters fall into the category of having 'excellent' intrarater reliability, these are arbitrary cut-off points. As JH's intrarater reliability is significantly lower than the rest of the group of raters, this rater will not be used in rating the videos for the Big CACTUS project. This is consistent with the fact that this rater had < 70% of scores within 0.5 between time points 1 and 2.

TABLE 26 Therapy outcome measures ratings of 10 videos by all raters: time point 1

Videos	Rater																		Median
	HW	EJ	AM	AF	VSm	JR	JC	JB	LM	HH	JH	MJ	SP	NU	JB	VSp	AP	SM	
1	3.5	3.5	3.5	4	4	4	4	3.5	3.5	4	4	4	4	3.5	3.5	4	4	3.5	4
2	4	4	4	4.5	4.5	4	4	4	4	4.5	4.5	4.5	4.5	4	4	3	3	4	4
3	1.5	2	2	2.5	2.5	1	1	2	1	1.5	1	2	1	1.5	1	1	1.5	2	1.5
4	2.5	2.5	2.5	3	3	1.5	1.5	2.5	1.5	2	2.5	2.5	1.5	1.5	1.5	1.5	2	2	2
5	1.5	2	1.5	1.5	2	1	1.5	1.5	1.5	1.5	1	2	1	1.5	1	0.5	1	1.5	1.5
6	2.5	2.5	2	2	2.5	1.5	2	2	2	1.5	2	3	1	1.5	1.5	1	2	2.5	2
7	3	3	3	3	4	3.5	3	3	3	3.5	3	3.5	2	3	3	1.5	2	3.5	3
8	2	3	3	3	4	3	3	3	3	3	2.5	3.5	3	3	3	1.5	1	3.5	3
9	3	2.5	2	1.5	2.5	1	1	1.5	2	2.5	1.5	2	1	2	1.5	0.5	1	1.5	1.5
10	3	2.5	2.5	2.5	2.5	2	1.5	2	1.5	2.5	2.5	2.5	1	2.5	2	1	1.5	2	2.25

■ 2.0 below 0.5 from median score
 ■ 1.5 below 0.5 from median score
 ■ 1.0 below 0.5 from median score
 ■ 0.5 below 0.5 from median score
 ■ 0.5-0.75 above 0.5 from median score
 ■ 1.0 above 0.5 from median score

TABLE 27 Inter-rater reliability of raters at time point 2 (excluding raters with low inter-rater reliability at time point 1 and/or low intrarater reliability)

Videos	Rater																Median	
	HW	EJ	AM	AF	VSm	JR	JB	LM	HH	JH	MJ	SP	NU	JB	VSp	AP		SM
1	4	3.5	4	4	4	4.5	4	4	4	4	4	3.5	4	3.5	4	3	3.5	4
2	4.5	4	4.5	4	4.5	4	4.5	4.5	4.5	5	4.5	4	4	4	3	4	4	4
3	2	2.5	1	1.5	2.5	1	1.5	1.5	2	2.5	2.5	1	1.5	1.5	1	1	2	1.5
4	2.5	3	1.5	3	2.5	1	2.5	2	2.5	3	2.5	1	2	1.5	1.5	2	2.5	2.5
5	1	2.5	1.5	2	3	1.5	1.5	2	2	1.5	2.5	1	1.5	1	0.5	1	2	1.5
6	2	3	2	3	2	1	1	2.5	1.5	2	2	1	1	1	1	1.5	1.5	1.5
7	2	3.5	3.5	3.5	3.5	4.5	4	3	4	4	4	2.5	3.5	3.5	1.5	2	3.5	3.5
8	1.5	3	3.5	3	3.5	4	3.5	3.5	4	3.5	3.5	3.5	3	3	1.5	1	3	3.5
9	2.5	2.5	3	2	3	1.5	1.5	2.5	2.5	3	2.5	1.5	1.5	1.5	0.5	1.5	2	2
10	3	2.5	2.5	2	2.5	2	2	3	3	2.5	2	1	1.5	2	1	2	1.5	2

■ 2.0 below 0.5 from median score
 ■ 1.5 below 0.5 from median score
 ■ 1.0 below 0.5 from median score
 ■ 0.5 below 0.5 from median score
 ■ 0.5-0.75 above 0.5 from median score
 ■ 1.0 above 0.5 from median score

TABLE 28 Intrarater reliability for each rater in *Tables 26 and 27*: TOMs benchmarking, reliability scores and intrarater reliability scores

Indicators of reliability	Rater																
	HW	EJ	AM	AF	VSm	JR	JB	LM	HH	JH	MJ	SP	NU	JB	VSp	AP	SM
Inter-rater reliability																	
Number of scores within 0.5 of median out of 20 (T1 +T2)	14	16	18	17	13	17	19	17	18	19	17	14	20	19	7	14	20
Percentage of scores within 0.5 of median	70	80	90	85	65	85	95	85	90	95	85	70	100	95	35	70	100
Intrarater (test–retest) reliability																	
Number of scores within 0.5 between T1 and T2	9	10	7	8	9	8	8	9	9	6	9	10	9	10	10	8	9
Percentage of scores within 0.5 between T1 and T2	90	100	70	80	90	80	80	90	90	60	90	100	90	100	100	80	90
ICCs for intrarater reliability	0.85	0.83	0.80	0.81	0.83	0.92	0.85	0.81	0.91	0.75	0.86	0.96	0.90	0.95	1.0	0.83	0.86

ICC, intraclass correlation coefficient; T1, time point 1; T2, time point 2.

Interpretation of ICCs: < 0.40 = poor; 0.40–0.59 = fair; 0.60–0.74 = good; 0.75–1.00 = excellent.

Shading denotes scores that fall below the reliability cut-off point for inclusion of raters to score all of the videoed conversations.

Fourteen raters were selected to rate videos for Big CACTUS. In all, 86% of their ratings were within 0.5 of the median (inter-rater reliability) and 88% of the ratings were within 0.5 between time points 1 and 2.

Although we have selected the 14 raters who rate most consistently compared with the median, and between time points 1 and 2, some variability still exists between raters. To maximise the chance that a change in score between time points represents a change in communication ability rather than a slight difference between raters, each participant will have all four videos rated by the same rater. To ensure that any differences in raters are spread evenly across the three trial groups, participants from the same site will be allocated to the same rater so each rater scores participants from each trial group.

A slight upwards trend was observed between ratings at time points 1 and 2, suggesting that there may be a familiarity effect. To account for this possibility, the order of presentation of videos from each participant will be randomised (e.g. participant 1, pair 1: 6 months–baseline; participant 1, pair 2: 9–12 months).

In some instances, there are only baseline videos available (due to withdrawal, etc.). In the SAP, the ITT analysis will be conducted based on only participants for whom there is an outcome measure. However, all data will be used to conduct a sensitivity analysis; therefore, baseline-only videos will be rated, but this will be conducted after all available pairs of videos have been rated.

Procedure for Therapy Outcome Measures rating document

Procedure for TOMS rating



You will receive videos in pairs, e.g. R01-01 A1 and R01-01 B1 (both of the same participant)

- 1) Use a timer to watch only the first 10 minutes of the first video in the pair and score according to the activity scale of the TOMs for Aphasia
 - Start at the top or bottom of the TOMs scale based on your first impression of whether it is good/poor and work up or down through the descriptors.
 - Remember that you can select a score half way between two of the written descriptors (scoring .5)
 - Focus on the underlined part of the descriptors as these can be judged through watching a video
- 2) Use a timer to watch only the first 10 minutes of the second video of the pair and score according to the activity scale of the TOMs for Aphasia
 - Start at the top or bottom of the TOMs scale based on your first impression of whether it is good/poor and work up or down through the descriptors.
 - Remember that you can select a score half way between two of the written descriptors (scoring .5)
 - Focus on the underlined part of the descriptors as these can be judged through watching a video
- 3) Check you are happy with the scores you have allocated for both videos. Modify if needed to ensure that if one video seemed better than the other, that the scores reflect this (NB this is likely to be a small change to the score only)
- 4) Document the final scores for both videos before moving on to the next pair of videos

Prompting/cueing includes:

- Therapist checking what the patient meant if it was ambiguous
- Therapist giving cues (e.g. semantic or phonemic)
- Therapist providing lead in phrases

NOT summarising/recapping what has been said in the conversation

TOM descriptors with clarifications following benchmarking session

KEY QUESTIONS to help choose starting point:

- 1) How successful is the participant at getting their message across?
- 2) How much help do they need in order to do this?

0	Unable to communicate in any way. No effective communication. No interaction.
1	Occasionally able to make *basic needs known with familiar persons or trained listener in familiar contexts. Minimal communication with maximal assistance. *Basic needs = everyday needs such as needing the toilet/ a drink
2	Limited functional communication. Consistently able to make *basic needs/conversation understood but is heavily dependent on cues and context. Communicates better with trained listener or family members or in familiar settings. Frequent repetition required. Maintains meaningful interaction *related to here and now. *Basic needs/conversation = everyday needs/simple conversation *Related to here and now = things that are in the immediate physical environment and can be pointed to/shown to the listener
3	Consistently able to *make needs known but can sometimes convey more information than this. Some inconsistency in unfamiliar settings. Is less dependent for intelligibility on cues and context. Occasional repetition required. *Communicates beyond here/now with *familiar persons; needs cues and prompting. *Make needs known = or able to engage in simple conversation *Communicates beyond here/now = talking about things are outside the immediate physical environment i.e. things that can be seen/heard from where the conversation is had. (This would include talking about things in the past/future but would not be limited to this). *Familiar person = or trained listener
4	Can be understood most of the time by any listener despite communication irregularities. Holds conversation; requires occasional prompts, particularly with a wider range of people.
5	Communicates effectively in all situations

*Half points can be allocated

Appendix 3 Definitions of adverse events and serious adverse events and categories of serious adverse event results

Safety assessments

Adverse events associated with the intervention are not anticipated given the low-risk intervention (in line with similar studies managed by Sheffield CTRU). However, if adverse events do occur these will be recorded by the therapist on the case report form and database. Adverse events do not need to be reported by fax to the CTRU.

Adverse events may include increased fatigue, fits or seizures, worsening vision or visual difficulties, increasing frequency or severity of headaches, accidents (e.g. falls) or injuries.

If a hospital admission or any other event considered serious occurs, these will be reported as serious adverse events (SAEs). We will not report further stroke-related events as SAEs as these are expected within this population.

The following criteria will be used when assessing SAEs.

Intensity (severity):

- mild – does not interfere with routine activities
- moderate – interferes with routine activities
- severe – impossible to perform routine activities.

Relationship to the trial activity (computerised speech therapy or puzzle books):

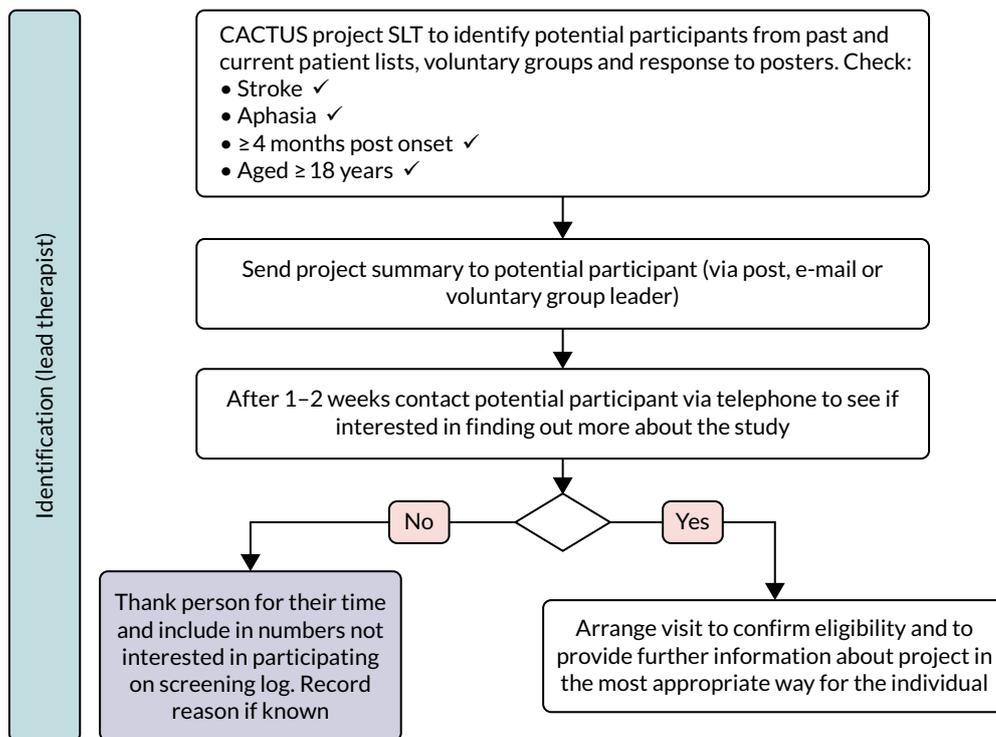
- Unrelated – there is no evidence of any causal relationship.
- Unlikely – there is little evidence to suggest that there is a causal relationship. There is another reasonable explanation for the event (e.g. the participant's clinical condition).
- Possible – there is some evidence to suggest a causal relationship. However, the influence of other factors may have contributed to the event (e.g. the participant's clinical condition).
- Probable – there is evidence to suggest a causal relationship and the influence of other factors is unlikely.
- Definite – there is clear evidence to suggest a causal relationship and other possible contributing factors can be ruled out.
- Not assessable – there is insufficient or contradictory information that cannot be supplemented or verified.

TABLE 29 Category of SAE (treatment as received), unrelated or unlikely to be related to trial activity

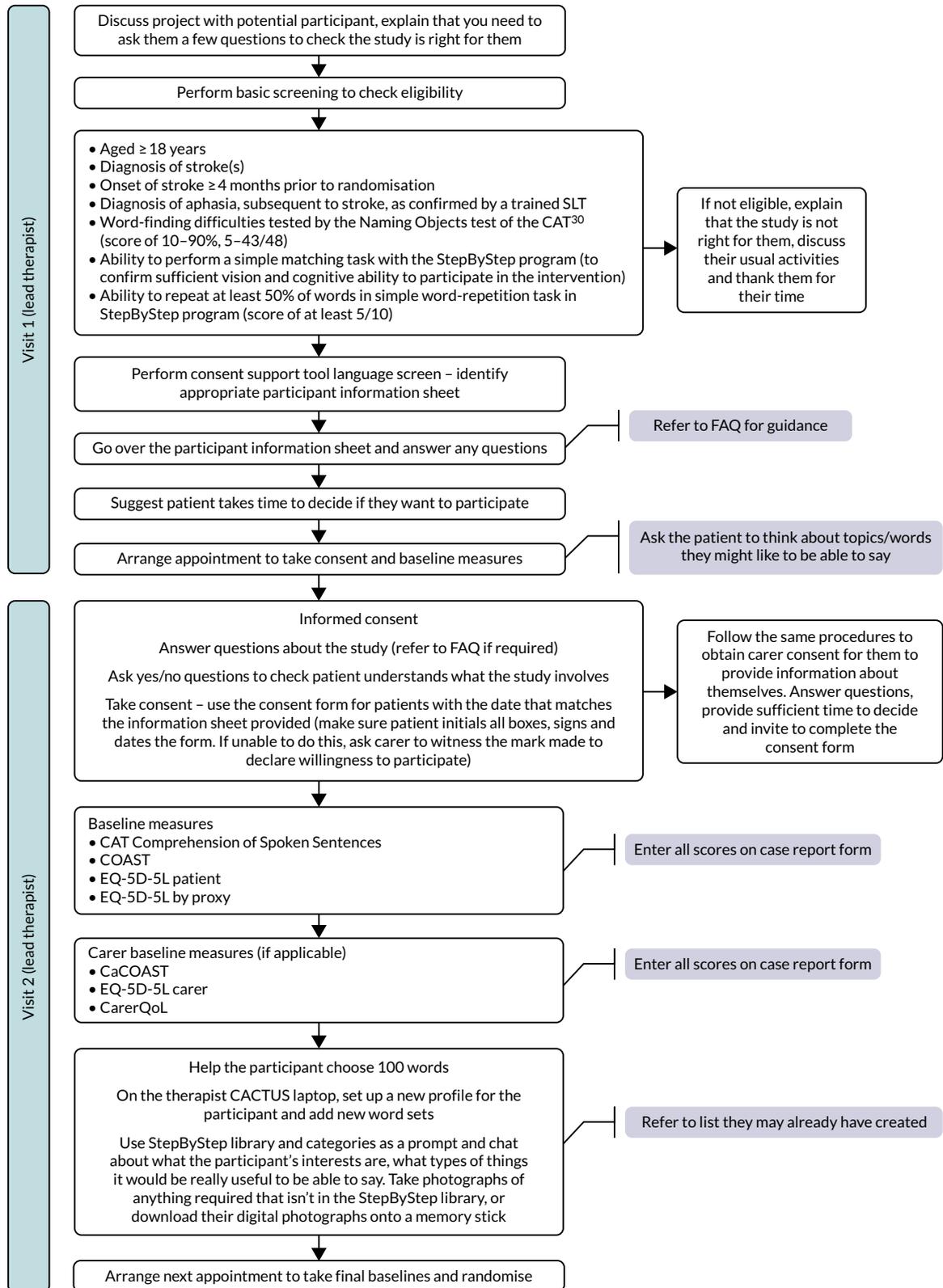
SAE category	Trial group, number of events		
	UC (n = 23)	AC (n = 12)	CSLT (n = 10)
Abdominal pain	0	1	0
Admission, cause unknown	0	0	2
Cardiac problems	0	1	1
Chest infection	1	0	1
Death, cardiac arrest	2	0	1
Death, cause unknown	1	1	0
Death, illness	1	0	0
Death, influenza	0	0	1
Death, sepsis	1	0	0
Diarrhoea and vomiting	2	0	0
Fall	4	0	1
Fall, fracture	3	3	1
Fracture	0	1	0
Infection	2	0	0
Muscular chest pain	0	1	0
Overdose	2	0	0
Possible seizure	0	1	0
Rectal bleeding	1	0	0
Seizure	0	0	1
Seizures, urine infection	1	0	0
Septicaemia	0	2	0
Urinary problems	1	1	0
Urinary tract infection	1	0	1

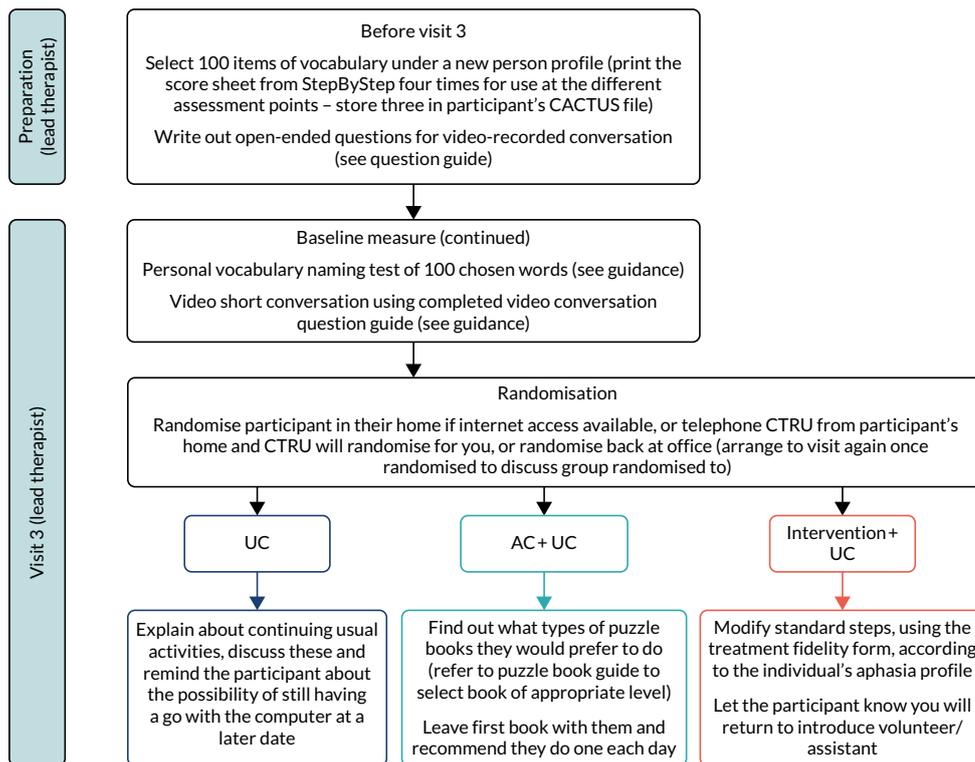
Appendix 4 Big CACTUS flow diagrams of activity

Identification

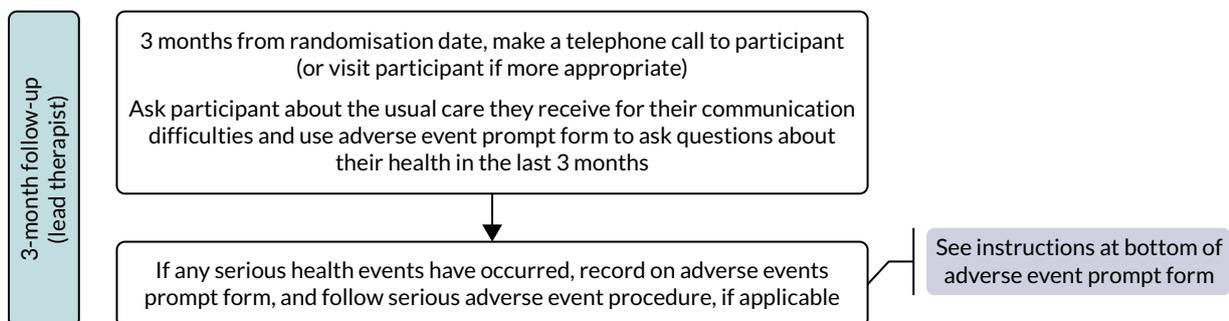


The baseline assessment and randomisation (guidance for taking consent and baseline information for CACTUS participants)

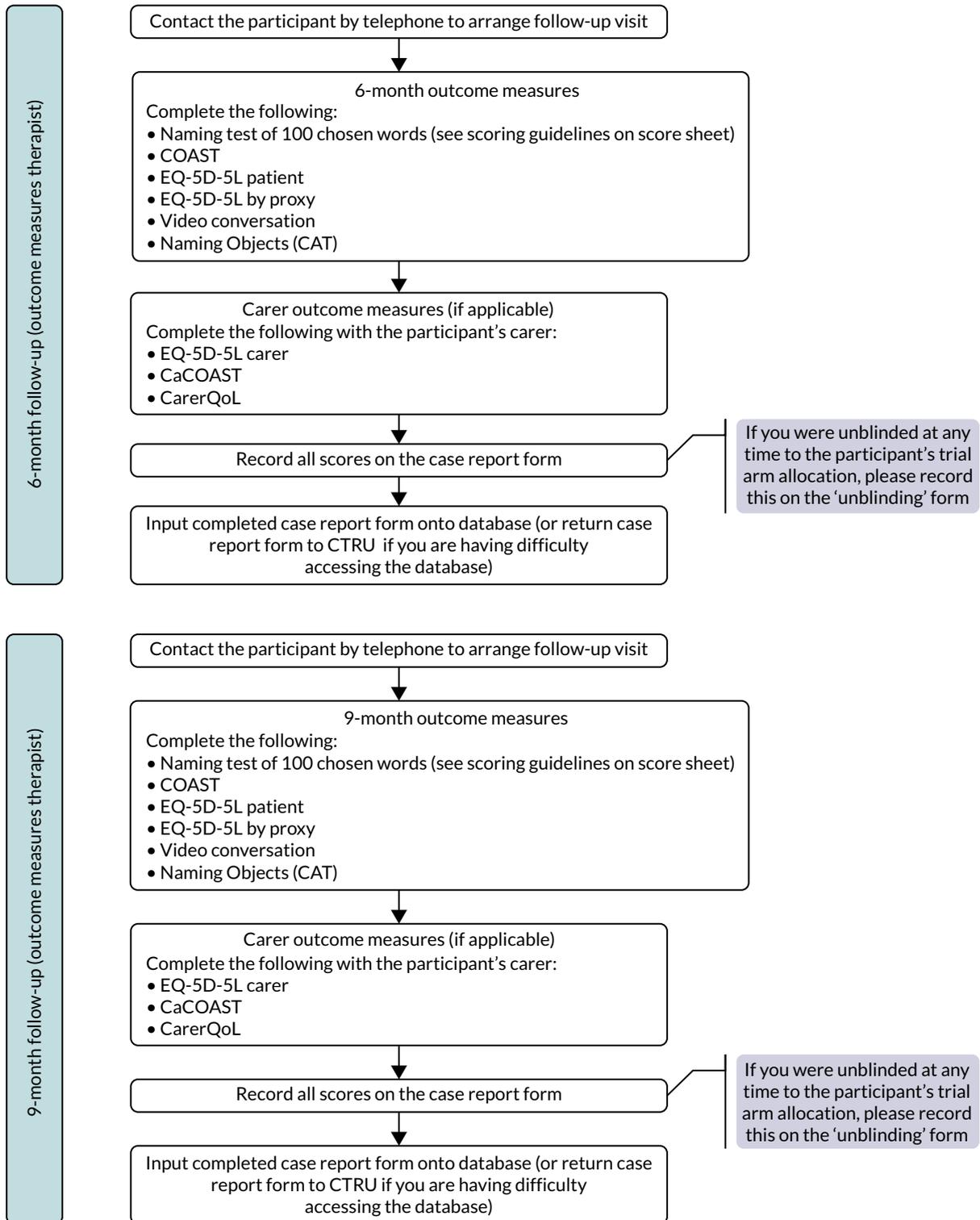


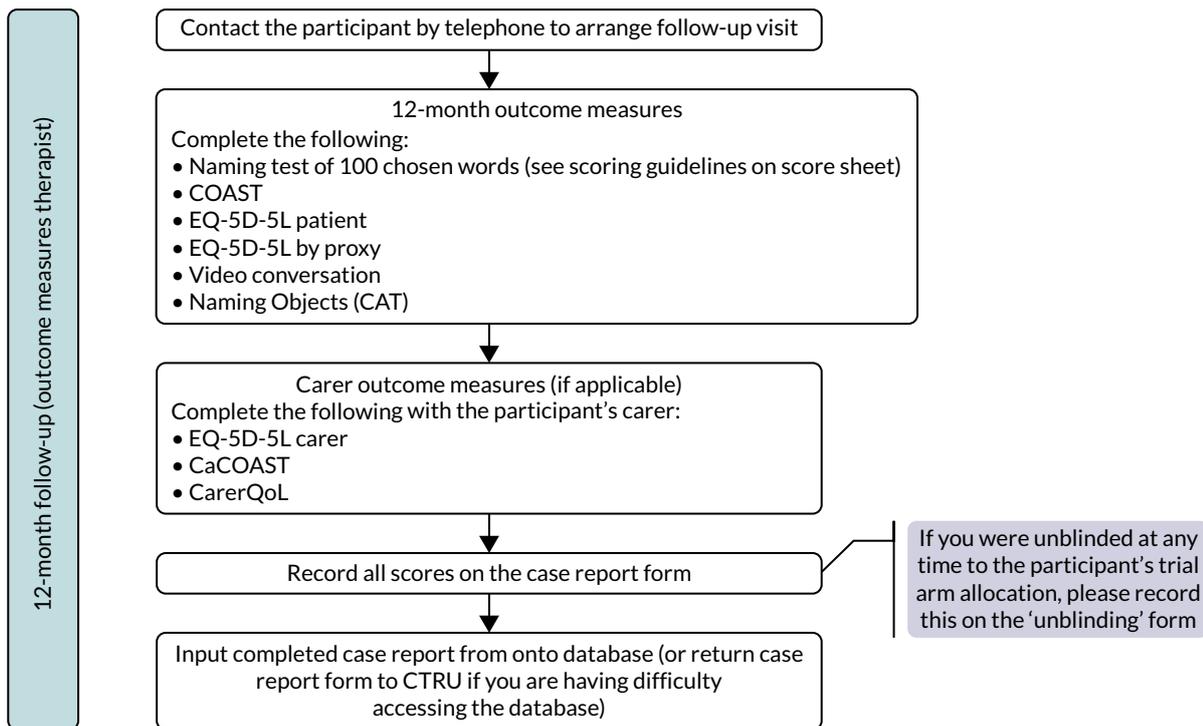


The 3-month follow-up assessments



The 6-, 9- and 12-month follow-up assessments





Appendix 5 Computer practice

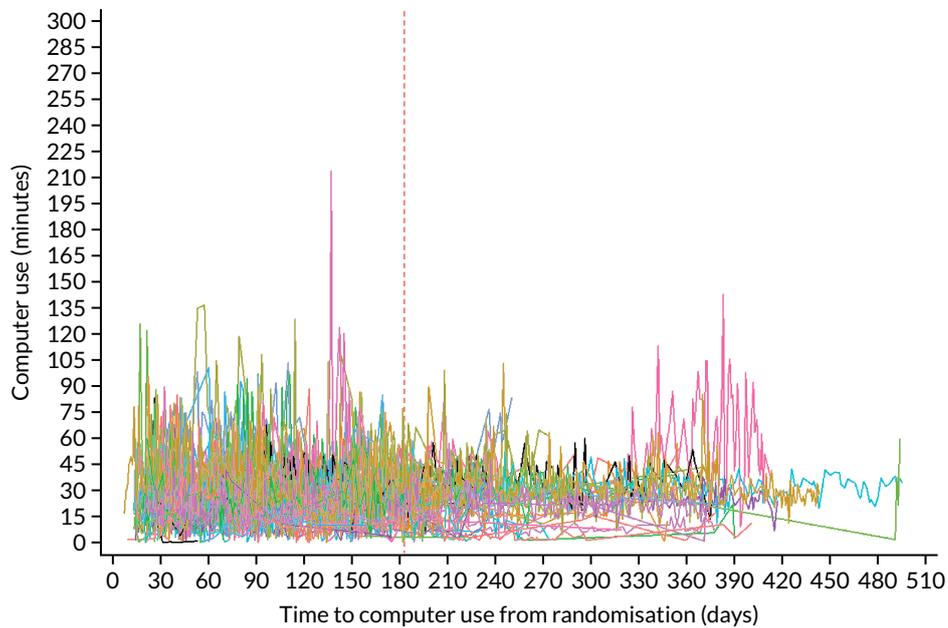


FIGURE 26 Computer therapy practice time per participant over time. Each line, regardless of colour, represents the participant's computer practice activity. The vertical dotted line is the expected 6-month assessment from randomisation.

Appendix 6 Unit costs

TABLE 30 Unit costs applied for valuation of resource use

Item description	Unit cost (£)	Source of unit cost	Note
Laptop/tablet loan for 6 months (for participants who did not have their own computer)	69.00	Big CACTUS trial	Unit cost calculated from the average cost of a laptop/tablet purchased through the NHS (£690) and divided by 10 users over its shelf life
StepByStep software individual licence	250.00	StepByStep website ²⁰	
StepByStep software clinician licence	550.00	StepByStep website ²⁰	
StepByStep software clinician five-licence bundle	2200.00	StepByStep website ²⁰	
Headsets	14.50	Big CACTUS trial	
Puzzle books	2.50	Big CACTUS trial	Average cost of a puzzle book purchased in the Big CACTUS trial
SLT band 7 cost per minute	0.90	PSSRU 2017 ⁶⁸	Delivery of training on StepByStep software
SLT band 6 cost per minute	0.75	PSSRU 2017 ⁶⁸	Delivery of CSLT intervention
SLT band 5 cost per minute	0.57	PSSRU 2017 ⁶⁸	Delivery of AC intervention
SLTA band 3 cost per minute	0.41	PSSRU 2017 ⁶⁸	Delivery of CSLT intervention
Volunteer cost per minute	0.41	PSSRU 2017 ⁶⁸	Only included in broader perspective; volunteers costed the same as a SLTA for providing an equivalent service
Travel cost per mile	0.45	GOV.UK ⁶⁹	

Appendix 7 EuroQol-5 Dimensions, five-level version completion information

Table 31 shows the proportions of complete data for the EQ-5D-5L questionnaires collected during Big CACTUS. Complete data for the EQ-5D-5L means that the patient completed all five items of the EQ-5D-5L questionnaire. Before assessments of data completeness were made, the EQ-5D-5L scores for the patients who died during the course of the trial were imputed as zero values after the patient's death. Note that the total numbers at baseline reported in Table 31 include the total number randomised to each trial group, including participants who subsequently died. Completion rates were high for the accessible EQ-5D-5L measure and therefore low levels of imputation were required for the base-case analysis.

TABLE 31 Data completeness for EQ-5D-5L questionnaires

Parameter	Trial group, n (%)		
	UC	AC	CSLT
EQ-5D-5L accessible	101 (100)	80 (100)	97 (100)
Baseline	101 (100)	80 (100)	96 (99)
6 months	89 (88)	71 (89)	84 (86)
9 months	86 (85)	64 (80)	80 (82)
12 months	87 (86)	62 (77)	77 (79)
EQ-5D-5L carer proxy	76 (75)	57 (71)	65 (65)
Baseline	76 (75)	57 (71)	65 (65)
6 months	53 (53)	44 (55)	57 (58)
9 months	47 (47)	35 (44)	51 (52)
12 months	49 (49)	34 (43)	45 (46)
EQ-5D-5L carer self-complete	62 (61)	49 (61)	64 (65)
Baseline	62 (61)	49 (61)	64 (65)
6 months	49 (48)	38 (47)	53 (55)
9 months	44 (44)	36 (45)	48 (50)
12 months	46 (45)	34 (42)	45 (46)

Appendix 8 Secondary analysis model parameters: utility scores

TABLE 32 Secondary analysis model parameters: utility scores (complete case, carer proxy, Hernandez-Alava *et al.*⁶³ mapping and English EQ-5D-5L tariff)

Parameter	Mean	Distribution (alpha, beta)	95% CI
Complete-case analysis			
Utility score (aphasia health state)	0.61	Beta (645.55, 405.25)	0.58 to 0.64
Difference in utility score associated with:		Distribution (standard error)	
Good response (6 months)	-0.04	Normal (0.02)	-0.09 to 0.01
Good response (9 months)	0.01	Normal (0.03)	-0.04 to 0.06
Good response (12 months)	0.03	Normal (0.03)	-0.02 to 0.07
Carer proxy analysis			
Utility score (aphasia health state)	0.51	Beta (642.11, 612.12)	0.48 to 0.54
Difference in utility score associated with:		Distribution (standard error)	
Good response (6 months)	-0.01	Normal (0.03)	-0.06 to 0.04
Good response (9 months)	-0.00	Normal (0.03)	-0.05 to 0.06
Good response (12 months)	0.02	Normal (0.03)	-0.04 to 0.08
Hernandez-Alava <i>et al.</i>⁶³ analysis			
Utility score (aphasia health state)	0.61	Beta (645.98, 421.65)	0.58 to 0.63
Difference in utility score associated with:		Distribution (standard error)	
Good response (6 months)	-0.04	Normal (0.02)	-0.09 to 0.00
Good response (9 months)	-0.02	Normal (0.03)	-0.07 to 0.03
Good response (12 months)	0.02	Normal (0.02)	-0.02 to 0.07
English EQ-5D-5L tariff analysis			
Utility score (aphasia health state)	0.71	Beta (895.15, 367.62)	0.68 to 0.73
Difference in utility score associated with:		Distribution (standard error)	
Good response (6 months)	-0.04	Normal (0.02)	-0.09 to 0.00
Good response (9 months)	-0.01	Normal (0.02)	-0.06 to 0.04
Good response (12 months)	0.02	Normal (0.02)	-0.03 to 0.06

Appendix 9 Computer and software costs scenario analysis

Table 33 presents the combined computer and software costs associated with the scenario analyses described in *Chapter 5, Model inputs: resource use and costs*. There is an appreciable change in costs under these different scenarios. It is important to note that when an individual lifetime licence is used it could be used beyond the 6-month intervention period (as was the case for participants with these licences in Big CACTUS). A clinician licence can be used by only one patient at a time and, therefore, would be taken away from a patient when it is transferred to another patient. Similarly, a 6-month individual licence would be usable for only a 6-month period. Therefore, the scenarios in *Table 33* that lead to reduced computer/software combined costs may also reduce the time for which a patient has access to the computer program, which may affect long-term outcomes.

TABLE 33 Computer and software costs scenario analyses

Scenario	Cost per patient (£)		
	Computer	Software	Combined computer and software
Base case (32% use own computer with an individual licence)	46.95	109.84	156.79
All CSLT participants borrow a computer and use a clinician StepByStep licence	69.00	44.00	113.00
All CSLT participants use their own computer and an individual StepByStep licence	0.00	250.00	250.00
75% of CSLT participants use their own computer and an individual StepByStep licence, 25% borrow a computer and use a clinician licence	17.25	198.50	215.75
£120 is paid for a 6-month individual StepByStep licence, rather than £250 for a lifetime licence, reflecting a new payment option	46.95	68.29	115.24

Appendix 10 Speech and language therapist, speech and language therapy assistant and volunteer costs

TABLE 34 The SLT, SLTA and volunteer costs

Description of cost	Time (hours per patient)	Unit cost (£) per hour (band)	Per-patient cost (£)
CSLT time and costs			
SLT time directly related to CSLT participants			
Setting up software	5.77	45.00 (band 6)	320.72
Providing technical support	0.99		
Monitoring progress	0.35		
Total	7.13		
SLT time related to support/training of SLTAs and volunteers			
Providing training	0.89	45.00 (band 6)	82.47
Supporting SLTA/volunteer	0.48		
Providing technical support	0.22		
Monitoring feedback form	0.24		
Total	1.83		
SLTA time directly related to CSLT participants			
Setting up computer/microphone	0.47	25.00 (band 3)	68.24
Encouraging the participant	0.98		
Assisting with software	0.58		
Conversations to practise words	0.51		
Other non-face-to-face contact	0.18		
Total	2.73		
SLTA time related to support/training			
Total time spent under supervision/ training from SLT	1.11	25.00 (band 3)	27.79
Volunteer time directly related to CSLT participants			
Setting up computer/microphone	0.11	25.00 (assumed equal to SLTA)	28.20
Encouraging the participant	0.50		
Assisting with software	0.30		
Conversations to practise words	0.20		
Other non-face-to-face contact	0.01		
Total	1.13		

continued

TABLE 34 The SLT, SLTA and volunteer costs (*continued*)

Description of cost	Time (hours per patient)	Unit cost (£) per hour (band)	Per-patient cost (£)
Volunteer time related to support/training			
Total time spent under supervision/ training from SLT	0.72	25.00 (assumed equal to SLTA)	18.03
AC time and costs			
Total time spent administering puzzle books	0.80	34.00 (band 5)	27.29
Total SLT/SLTA/volunteer costs			
Total SLT/SLTA costs associated with CSLT	12.80	-	499.22
Total volunteer costs associated with CSLT	1.85	-	46.22
Total SLT/SLTA costs associated with AC	0.80	-	27.29

Appendix 11 Model parameters: resource use

TABLE 35 Model parameters: resource use

Parameter	Mean	Distribution (parameters)	95% CI	Source
CSLT resource use parameters				
Proportion who need to borrow a computer	0.68	Beta (66, 31)		Big CACTUS
Utilisation rate of software licences	0.50	Beta (5, 5)		Assumption
Number of patients who could benefit from speech and language therapy StepByStep software training annually	55.43	Gamma (11.09)		Big CACTUS, with assumed standard error equal to 20% of the mean
Proportion who need a headset	0.33	Beta (32, 65)		Big CACTUS
Mean SLT time spent with participant (hours)	7.13	Gamma (0.54)	6.06 to 8.19	Big CACTUS
Mean SLT time spent supervising/training SLTAs and volunteers (hours)	1.83	Gamma (0.14)	1.55 to 2.10	Big CACTUS
Mean SLTA time spent with participants (hours)	2.73	Gamma (0.28)	2.17 to 3.28	Big CACTUS
Mean SLTA time spent under supervision of SLT (hours)	1.11	Gamma (0.12)	0.86 to 1.35	Big CACTUS
Mean SLT/SLTA mileage	127.52	Gamma (13.50)	100.67 to 154.35	Big CACTUS
Mean volunteer time spent with participants (hours)	1.13	Gamma (0.20)	0.72 to 1.52	Big CACTUS
Mean volunteer time spent under supervision of SLT (hours)	0.72	Gamma (0.14)	0.44 to 1.01	Big CACTUS
Mean volunteer mileage	23.57	Gamma (6.70)	10.29 to 36.83	Big CACTUS
AC resource use parameters				
Mean number of puzzle books	4.35	Gamma (0.20)	3.96 to 4.74	Big CACTUS
Mean SLT time spent administering AC (hours)	0.80	Gamma (0.05)	0.70 to 0.89	Big CACTUS
Parameters for beta distribution: alpha, beta. Parameter for gamma distribution: standard error.				

Appendix 12 Model parameters used for base-case, secondary and subgroup analyses

TABLE 36 Model parameters used for base-case, secondary and subgroup analyses

Model parameter	Distribution	Distribution parameters	Values in each analysis modelled																
			Base case	Stroke time quartile SGs				Comprehension ability SGs				Word-finding ability SGs			Secondary analyses				
				Q1	Q2	Q3	Q4	Mild	Moderate	Severe	Norm	Mild	Moderate	Severe	1. Complete-case analysis	2. English EQ-5D-5L tariff	3. Proxy report	4. EQ-5D-5L mapping of Hernandez-Alava et al. ⁶³	5. Including carer QALYs
6-month good response: CSLT	Beta	Alpha	73	21	17	22	13	31	24	4	14	30	26	19	73	73	73	73	73
		Beta	21	6	4	4	7	12	5	1	3	11	4	6	21	21	21	21	21
6-month good response: UC	Beta	Alpha	45	11	13	12	9	24	7	3	11	19	18	10	45	45	45	45	45
		Beta	52	11	14	12	15	26	17	0	9	21	19	14	52	52	52	52	52
6-month good response: AC	Beta	Alpha	39	10	9	9	11	21	11	2	5	19	9	12	39	39	39	39	39
		Beta	40	8	11	7	14	15	13	4	8	19	13	10	40	40	40	40	40
9-month new response: CSLT	Beta	Alpha	7	2	4	2	3	4	3	0	0	3	1	3	7	7	7	7	7
		Beta	14	4	0	2	4	8	2	1	3	8	3	3	14	14	14	14	14
9-month new response: UC	Beta	Alpha	19	4	3	5	7	10	5	0	4	9	9	2	19	19	19	19	19
		Beta	33	7	11	7	8	16	12	0	5	12	10	12	33	33	33	33	33
9-month new response: AC	Beta	Alpha	14	6	4	2	2	3	6	2	3	7	3	4	14	14	14	14	14
		Beta	26	2	7	5	12	12	7	2	5	12	10	6	26	26	26	26	26
9-month relapse: CSLT	Beta	Alpha	14	5	4	3	2	5	8	1	0	4	5	5	14	14	14	14	14
		Beta	59	16	13	19	11	26	16	3	14	26	21	14	59	59	59	59	59
9-month relapse: UC	Beta	Alpha	10	3	1	4	2	5	1	0	4	7	3	1	10	10	10	10	10
		Beta	35	8	12	8	7	19	6	3	7	12	15	9	35	35	35	35	35
9-month relapse: AC	Beta	Alpha	11	2	1	4	4	5	5	1	0	3	2	6	11	11	11	11	11
		Beta	28	8	8	5	7	16	6	1	5	16	7	6	28	28	28	28	28
12-month new response: CSLT	Beta	Alpha	11	2	4	3	2	5	6	0	0	3	4	4	11	11	11	11	11
		Beta	17	7	4	2	4	8	4	2	3	9	4	4	17	17	17	17	17
12-month new response: UC	Beta	Alpha	16	4	3	7	2	8	4	0	4	8	2	6	16	16	16	16	16
		Beta	27	6	9	4	8	13	9	0	5	11	11	7	27	27	27	27	27
12-month new response: AC	Beta	Alpha	7	1	1	2	3	4	3	0	0	3	1	3	7	7	7	7	7
		Beta	30	3	7	7	13	13	9	3	5	12	11	9	30	30	30	30	30

Model parameter	Distribution	Distribution parameters	Values in each analysis modelled																
			Base case	Stroke time quartile SGs				Comprehension ability SGs				Word-finding ability SGs			Secondary analyses				
				Q1	Q2	Q3	Q4	Mild	Moderate	Severe	Norm	Mild	Moderate	Severe	1. Complete-case analysis	2. English EQ-5D-5L tariff	3. Proxy report	4. EQ-5D-5L mapping of Hernandez-Alava <i>et al.</i> ⁶³	5. Including carer QALYs
12-month relapse: CSLT	Beta	Alpha	5	0	0	3	2	3	2	0	0	2	1	2	5	5	5	5	5
		Beta	61	18	13	18	12	27	17	3	14	27	21	15	61	61	61	61	61
12-month relapse: UC	Beta	Alpha	12	5	1	1	5	8	3	1	0	5	4	3	12	12	12	12	12
		Beta	42	7	14	12	9	21	8	2	11	16	20	8	42	42	42	42	42
12-month relapse: AC	Beta	Alpha	8	3	4	0	1	4	3	1	0	3	4	1	8	8	8	8	8
		Beta	34	11	8	7	8	15	9	2	8	20	6	9	34	34	34	34	34
Baseline EQ-5D-5L	Normal	Mean	0.61	0.64	0.65	0.61	0.56	0.64	0.6	0.58	0.58	0.62	0.63	0.59	0.61	0.71	0.51	0.61	0.77
		Standard error	0.01	0.03	0.03	0.03	0.03	0.02	0.03	0.09	0.04	0.02	0.03	0.03	0.02	0.01	0.01	0.01	0.02
6-month EQ-5D-5L difference	Normal	Mean	-0.04	-0.07	-0.04	-0.03	-0.01	0.03	-0.08	-0.18	-0.1	-0.03	-0.03	-0.04	-0.04	-0.04	-0.01	-0.04	-0.01
		Standard error	0.03	0.07	0.05	0.05	0.05	0.04	0.05	0.15	0.07	0.04	0.05	0.06	0.03	0.02	0.03	0.02	0.03
9-month EQ-5D-5L difference	Normal	Mean	-0.02	-0.12	-0.03	-0.06	0.1	0.01	-0.01	-0.24	-0.07	-0.04	0.08	-0.06	0.01	-0.01	0	-0.02	0.01
		Standard error	0.03	0.06	0.05	0.05	0.06	0.04	0.05	0.15	0.07	0.04	0.04	0.06	0.03	0.02	0.03	0.02	0.03
12-month EQ-5D-5L difference	Normal	Mean	0.02	0.01	0.07	-0.1	0.06	-0.01	0.08	-0.1	0.02	0.03	0.02	-0.01	0.03	0.02	0.02	0.02	-0.02
		Standard error	0.03	0.06	0.04	0.06	0.05	0.04	0.04	0.14	0.06	0.04	0.04	0.06	0.03	0.02	0.03	0.02	0.03
SLT with participant (mean number of hours)	Gamma	Mean	7.13	7.48	7.75	6.83	6.34	6.39	8	10.88	6.5	5.73	8.6	8.03	7.13	7.13	7.13	7.13	7.13
		Standard error	0.54	0.83	1.47	0.99	1.09	0.77	1.14	3.37	0.77	0.62	1.12	1.09	0.54	0.54	0.54	0.54	0.54
SLT with SLTA/volunteer (mean number of hours)	Gamma	Mean	1.83	1.46	1.79	2.3	1.76	1.67	1.99	2.72	1.75	1.76	1.82	1.95	1.83	1.83	1.83	1.83	1.83
		Standard error	0.14	0.22	0.23	0.34	0.25	0.23	0.24	0.76	0.22	0.22	0.23	0.26	0.14	0.14	0.14	0.14	0.14
SLTA with participant (mean number of hours)	Gamma	Mean	2.73	2.48	2.92	3.02	2.49	2.26	3.53	1.92	2.79	2.48	3.48	2.6	2.73	2.73	2.73	2.73	2.73
		Standard error	0.28	0.5	0.63	0.51	0.65	0.39	0.56	0.92	0.63	0.4	0.54	0.57	0.28	0.28	0.28	0.28	0.28

continued

TABLE 36 Model parameters used for base-case, secondary and subgroup analyses (continued)

Model parameter	Distribution	Distribution parameters	Values in each analysis modelled																
			Base case	Stroke time quartile SGs				Comprehension ability SGs				Word-finding ability SGs			Secondary analyses				
				Q1	Q2	Q3	Q4	Mild	Moderate	Severe	Norm	Mild	Moderate	Severe	1. Complete-case analysis	2. English EQ-5D-5L tariff	3. Proxy report	4. EQ-5D-5L mapping of Hernandez-Alava <i>et al.</i> ⁶³	5. Including carer QALYs
SLTA with SLT (mean number of hours)	Gamma	Mean	1.11	0.96	1.13	1.27	1.08	0.98	1.35	1.83	0.86	0.89	1.38	1.21	1.11	1.11	1.11	1.11	1.11
		Standard error	0.12	0.23	0.26	0.23	0.3	0.18	0.23	1.01	0.2	0.16	0.26	0.24	0.12	0.12	0.12	0.12	0.12
Volunteer with participant (mean number of hours)	Gamma	Mean	1.13	0.84	1.19	1.13	1.45	1.12	1.13	1.08	1.16	1.1	0.97	1.27	1.13	1.13	1.13	1.13	1.13
		Standard error	0.2	0.29	0.47	0.41	0.49	0.31	0.41	0.7	0.39	0.29	0.37	0.43	0.2	0.2	0.2	0.2	0.2
Volunteer with SLT (mean number of hours)	Gamma	Mean	0.72	0.5	0.66	1.03	0.68	0.68	0.63	0.88	0.89	0.87	0.43	0.74	0.72	0.72	0.72	0.72	0.72
		Standard error	0.14	0.18	0.24	0.41	0.22	0.23	0.25	0.61	0.3	0.25	0.16	0.29	0.14	0.14	0.14	0.14	0.14
Number of puzzle books	Gamma	Mean	4.35	3.83	4.75	3.88	4.72	4.42	4.17	4.2	4.57	4.38	4.27	4.43	4.35	4.35	4.35	4.35	4.35
		Standard error	0.2	0.44	0.49	0.39	0.25	0.31	0.32	0.73	0.52	0.32	0.33	0.31	0.2	0.2	0.2	0.2	0.2
SLT administering AC (mean hours)	Gamma	Mean	0.8	0.73	0.95	0.78	0.75	0.77	0.86	0.7	0.84	0.86	0.8	0.71	0.8	0.8	0.8	0.8	0.8
		Standard error	0.05	0.11	0.1	0.1	0.07	0.08	0.09	0.17	0.09	0.07	0.09	0.08	0.05	0.05	0.05	0.05	0.05

OHE, Office of Health Economics; Q, quartile; SG, subgroup.

Note

For beta distributions, alpha = number of people with the event and beta = number without the event.

Appendix 13 Expected value of information results

Based on EVPI analysis conducted on the base-case model analysis, we estimate that the per-patient value of perfect information is £160.55 given a cost-effectiveness threshold of £20,000 per QALY gained. We extrapolated the per-patient value of information estimates to a population level by estimating the number of participants who would be likely to receive the treatment over a 10-year period. It is estimated that the prevalence of aphasia in the UK is 350,000.⁹⁴ Thirty-three per cent of people with aphasia screened for potential recruitment into Big CACTUS were not eligible and, conservatively, we assume that the intervention would not be offered to people with severe word-finding difficulty (26% of the Big CACTUS population). Hence, we estimate that 172,640 people in the UK would be eligible for CSLT treatment ($350,000 \times 0.67 \times 0.74$). It is estimated that the incidence of aphasia is 11,400 per year in Great Britain.⁹⁵ Twenty-four per cent of these people recover in the first 6 months, but 76% do not.⁹⁵ Given the eligibility for CSLT treatment described above, we estimate that the annual incidence population who would be eligible for CSLT treatment is 4274 ($11,400 \times 0.76 \times 0.67 \times 0.74$). Over a 10-year period, we therefore estimate that 215,376 people could be treated with CSLT in the UK, with an average of 21,538 treated per year. Based on this, we estimate a population-level EVPI of £34.6M. The overall EVPI per patient for a range of cost-effectiveness thresholds is presented in *Figure 27*.

A further analysis was undertaken to gain an insight into which areas would deliver the most value for further research. The Sheffield Accelerated Value of Information tool⁷² was used to estimate the value of obtaining perfect information for one parameter or group of parameters. The following parameters were included in the analysis:

- utility score in aphasia health state
- utility difference associated with response in each of the response health states (6, 9, 12 months and beyond)
- probability of good response at each time point for CSLT (6, 9 and 12 months)
- probability of good response at each time point for AC (6, 9 and 12 months)
- probability of good response at each time point for UC (6, 9 and 12 months)
- probability of relapse at each time point for CSLT (6–9 months, 9–12 months and beyond 12 months)
- probability of relapse at each time point for AC (6–9 months, 9–12 months and beyond 12 months)
- probability of relapse at each time point for UC (6–9 months, 9–12 months and beyond 12 months)
- proportion who need to borrow a computer
- software cost
- total SLT and SLTA costs for CSLT
- total SLT costs for AC
- mortality rate.

Only four parameters resulted in per-patient EVPI of > £0.03: utility difference associated with response at 12 months, and the relapse rates from 9 months onwards for each of the three treatment options. The expected value of partial perfect information for each of these parameters is presented in *Table 37*.

Clearly, the utility difference associated with a good response at 12 months is the most important parameter in the model, and obtaining more information on this would be highly valuable. This is unsurprising because this utility score is extrapolated into the future; it is primarily this score that dictates the QALY benefit associated with increasing the response rate. The importance of the longer-term relapse rates is also expected because it is these that dictate at what rate people return to the aphasia health state from the good response health state.

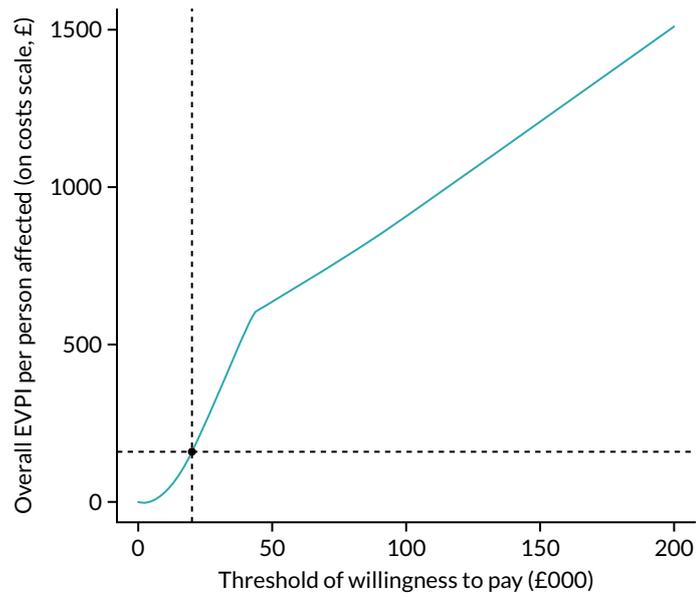


FIGURE 27 Overall EVPI (per patient).

TABLE 37 Expected value of partial perfect information: base-case analysis

Parameter	Per person EVPPI (£)	Standard error (£)	EVPPi for UK over 10 years (£)
Utility difference associated with response at 12 months	94.66	1.82	20,390,000
Probability of relapse from 9 months onwards for CSLT	5.20	1.26	1,119,000
Probability of relapse from 9 months onwards for AC	16.36	0.70	3,525,000
Probability of relapse from 9 months onwards for UC	7.64	0.80	1,645,000

EVPPi, expected value of partial perfect information.

Appendix 14 Results of model-based secondary analyses

Tables 38–40 present results from the 11 secondary analyses undertaken, which are described in Chapter 5, *Model results: secondary analyses*. Note that results for analyses 6 to 11 are not included in Table 40, which presents cost-effectiveness results for AC compared with UC, because these scenarios made no changes to the UC and AC treatment options.

TABLE 38 Cost-effectiveness results from base-case and secondary analyses: CSLT vs. UC model-based analysis

Analysis	CSLT cost (£)	UC cost (£)	Incremental cost (£): CSLT vs. UC (95% CrI)	CSLT QALYs	UC QALYs	Incremental QALYs: CSLT vs. UC (95% CrI)	ICER (£ per QALY gained)
Base-case analysis (cross-walk)	732.73	0.00	732.73 (674.23 to 798.05)	4.2164	4.1992	0.0172 (-0.05 to 0.10)	42,686
1. Complete-case analysis	733.08	0.00	733.08 (672.13 to 800.63)	4.2302	4.2042	0.0260 (-0.03 to 0.11)	28,248
2. Using English EQ-5D-5L tariff	732.25	0.00	732.25 (673.19 to 797.84)	4.8537	4.8406	0.0132 (-0.04 to 0.09)	55,639
3. Using carer proxy EQ-5D-5L measure	733.06	0.00	733.06 (672.70 to 800.01)	3.5339	3.5084	0.0254 (-0.05 to 0.12)	28,819
4. Use EQ-5D-5L mapping of Hernandez-Alava <i>et al.</i> ⁶³	732.96	0.00	732.96 (672.60 to 798.22)	4.1568	4.1358	0.0210 (-0.04 to 0.11)	34,921
5. Including carer QALYs	732.17	0.00	732.17 (673.70 to 798.05)	7.9251	7.9275	-0.0025 (-0.10 to 0.10)	Dominated
6. Assume that all CSLT participants loaned computer and use clinician licence	690.85	0.00	690.85 (632.45 to 766.49)	4.2133	4.1958	0.0175 (-0.04 to 0.10)	39,480
7. Assume that all CSLT participants use own computer and individual licence	822.16	0.00	822.16 (771.69 to 876.82)	4.2165	4.1985	0.0180 (-0.04 to 0.10)	45,813
8. Assume that 75% of CSLT participants use own computer	789.12	0.00	788.80 (736.66 to 844.44)	4.2187	4.2014	0.0173 (-0.04 to 0.10)	45,552
9. Pay £120 for a 6-month licence instead of £250 for lifetime licence	691.27	0.00	691.27 (635.30 to 754.07)	4.2135	4.1962	0.0173 (-0.05 to 0.10)	39,844
10. Zero cost of software	619.35	0.00	619.35 (NC)	4.2164	4.1992	0.0172 (-0.05 to 0.10)	36,081

continued

TABLE 38 Cost-effectiveness results from base-case and secondary analyses: CSLT vs. UC model-based analysis (continued)

Analysis	CSLT cost (£)	UC cost (£)	Incremental cost (£): CSLT vs. UC (95% CrI)	CSLT QALYs	UC QALYs	Incremental QALYs: CSLT vs. UC (95% CrI)	ICER (£ per QALY gained)
11. Zero SLT and SLTA costs	165.11	0.00	165.11 (NC)	4.2164	4.1992	0.0172 (-0.05 to 0.10)	9619
12. SLT and SLTA costs halved	448.92	0.00	448.92 (NC)	4.2164	4.1992	0.0172 (-0.05 to 0.10)	26,153
13. Assume same outcomes with 28 hours of band 6 face-to-face SLT time and travel	1342.78	0.00	1342.78 (NC)	4.2164	4.1992	0.0172 (-0.05 to 0.10)	78,068
14. Broader perspective	789.39	0.00	789.39 (728.09 to 858.03)	4.2166	4.1988	0.0178 (-0.05 to 0.10)	44,279
NC, not calculated.							

TABLE 39 Cost-effectiveness results from base-case and secondary analyses: CSLT vs. AC model-based analysis

Analysis	CSLT cost (£)	AC cost (£)	Incremental cost (£): CSLT vs. AC (95% CrI)	CSLT QALYs	AC QALYs	Incremental QALYs: CSLT vs. AC (95% CrI)	ICER (£ per QALY gained)
Base-case analysis (cross-walk)	732.73	38.14	694.59 (636.46 to 760.09)	4.2164	4.1991	0.0173 (-0.05 to 0.10)	40,164
1. Complete-case analysis	733.08	38.18	694.90 (633.87 to 762.28)	4.2302	4.2040	0.0262 (-0.03 to 0.11)	26,555
2. Using English EQ-5D-5L tariff	732.25	38.17	694.09 (634.95 to 759.75)	4.8537	4.8402	0.0135 (-0.04 to 0.09)	51,308
3. Using carer proxy EQ-5D-5L measure	733.06	38.18	694.88 (634.58 to 761.87)	3.5339	3.5085	0.0254 (-0.05 to 0.12)	27,397
4. Use EQ-5D-5L mapping of Hernandez-Alava <i>et al.</i> ⁶³	732.96	38.18	694.78 (634.94 to 760.21)	4.1568	4.1356	0.0211 (-0.04 to 0.11)	32,835
5. Including carer QALYs	732.17	38.14	694.03 (635.13 to 759.61)	7.9251	7.9271	-0.0020 (-0.10 to 0.10)	Dominated
6. Assume that all CSLT participants loaned computer and use clinician licence	690.85	38.20	652.65 (594.21 to 727.91)	4.2133	4.1938	0.0176 (-0.04 to 0.10)	37,091
7. Assume that all CSLT participants use own computer and individual licence	822.16	38.15	784.01 (732.89 to 838.53)	4.2165	4.1983	0.0182 (-0.04 to 0.10)	43,113
8. Assume that 75% of CSLT participants use own computer	789.12	38.18	750.95 (698.85 to 806.41)	4.2187	4.2013	0.0174 (-0.04 to 0.10)	43,051

TABLE 39 Cost-effectiveness results from base-case and secondary analyses: CSLT vs. AC model-based analysis (continued)

Analysis	CSLT cost (£)	AC cost (£)	Incremental cost (£): CSLT vs. AC (95% CrI)	CSLT QALYs	AC QALYs	Incremental QALYs: CSLT vs. AC (95% CrI)	ICER (£ per QALY gained)
9. Pay £120 for a 6-month licence instead of £250 for lifetime licence	691.27	38.19	653.08 (597.08 to 715.97)	4.2135	4.1961	0.0175 (-0.04 to 0.11)	37,346
10. Zero cost of software	619.35	38.19	581.21 (NC)	4.2164	4.1991	0.0173 (-0.05 to 0.10)	33,608
11. Zero SLT and SLTA costs	165.11	38.19	126.97 (NC)	4.2164	4.1991	0.0173 (-0.05 to 0.10)	7342
12. SLT and SLTA costs halved	448.92	38.19	410.78 (NC)	4.2164	4.1991	0.0173 (-0.05 to 0.10)	23,753
13. Assume same outcomes with 28 hours of band 6 face-to-face SLT time and travel	1342.78	38.19	1304.59 (NC)	4.2164	4.1991	0.0173 (-0.05 to 0.10)	75,049
14. Broader perspective	789.39	38.18	751.21 (689.57 to 819.71)	4.2166	4.1987	0.0179 (-0.05 to 0.10)	41,974
NC, not calculated.							

TABLE 40 Cost-effectiveness results from base-case and secondary analyses: AC vs. UC model-based analysis

Analysis	AC cost (£)	UC cost (£)	Incremental cost (£): AC vs. UC (95% CrI)	AC QALYs	UC QALYs	Incremental QALYs: AC vs. UC (95% CrI)	ICER (£ per QALY gained)
Base-case analysis (cross-walk)	38.14	0.00	38.14 (34.94 to 41.50)	4.1991	4.1992	-0.0001 (-0.02 to 0.02)	Dominated
1. Complete-case analysis	38.18	0.00	38.18 (34.89 to 41.66)	4.2040	4.2042	-0.0002 (-0.02 to 0.02)	Dominated
2. Using English EQ-5D-5L tariff	38.17	0.00	38.17 (34.92 to 41.53)	4.8402	4.8406	-0.0004 (-0.02 to 0.02)	Dominated
3. Using carer proxy EQ-5D-5L measure	38.18	0.00	38.18 (34.95 to 41.62)	3.5085	3.5084	0.0001 (-0.02 to 0.02)	522,118
4. Use EQ-5D-5L mapping of Hernandez-Alava <i>et al.</i> ⁶³	38.18	0.00	38.18 (34.90 to 41.57)	4.1356	4.1358	-0.0002 (-0.02 to 0.02)	Dominated
5. Including carer QALYs	38.14	0.00	38.14 (34.87 to 41.52)	5.2343	5.2348	-0.0005 (-0.03 to 0.02)	Dominated

Appendix 15 Model-based subgroup analysis results

TABLE 41 Results of subgroup analyses comparing CSLT with UC: model-based analysis

Analysis	Subgroup	CSLT cost (£)	UC cost (£)	Incremental cost (£): CSLT vs. UC (95% CrI)	CSLT QALYs	UC QALYs	Incremental QALYs: CSLT vs. UC (95% CrI)	ICER (£ per QALY gained)
Word-finding difficulty (baseline)	Mild	653.49	0.00	653.49 (586.44 to 728.36)	4.2856	4.2564	0.0292 (-0.06 to 0.17)	22,371
	Moderate	822.77	0.00	822.77 (715.54 to 942.22)	4.3290	4.3006	0.0284 (-0.10 to 0.22)	28,898
	Severe	778.71	0.00	778.71 (674.77 to 890.79)	3.9971	4.0137	-0.0166 (-0.16 to 0.11)	Dominated
Comprehension ability (baseline)	Normal	695.77	0.00	695.77 (613.47 to 785.33)	4.0083	4.0108	-0.0024 (-0.28 to 0.32)	Dominated
	Mild	677.12	0.00	677.12 (599.44 to 763.04)	4.3544	4.3595	-0.0051 (-0.10 to 0.08)	Dominated
	Moderate	804.44	0.00	804.44 (699.07 to 922.11)	4.1968	4.1360	0.0608 (-0.05 to 0.28)	13,235
	Severe	940.81	0.00	940.81 (669.69 to 1292.58)	3.6722	3.7654	-0.0932 (-0.95 to 0.42)	Dominated
Time post stroke (quartiles)	1 (shortest)	721.32	0.00	721.32 (637.43 to 816.21)	4.3279	4.3211	0.0068 (-0.36 to 0.39)	105,532
	2	763.36	0.00	763.36 (635.70 to 913.21)	4.6455	4.5172	0.1280 (-0.12 to 0.50)	5948
	3	752.31	0.00	752.31 (655.33 to 861.75)	4.0446	3.9979	0.04668 (-0.17 to 0.40)	16,115
	4 (longest)	687.08	0.00	687.08 (581.95 to 809.62)	3.9071	3.8632	0.0444 (-0.03 to 0.22)	15,663

TABLE 42 Results of subgroup analyses comparing CSLT with AC: model-based analysis

Analysis	Subgroup	CSLT cost (£)	AC cost (£)	Incremental cost (£): CSLT vs. AC (95% CrI)	CSLT QALYs	AC QALYs	Incremental QALYs: CSLT vs. AC (95% CrI)	ICER (£ per QALY gained)
Word-finding difficulty (baseline)	Mild	653.49	40.17	613.32 (545.64 to 689.09)	4.2856	4.2658	0.0198 (-0.07 to 0.16)	30,911
	Moderate	822.77	37.99	784.78 (677.99 to 904.96)	4.3290	4.2874	0.0416 (-0.12 to 0.26)	18,855
	Severe	778.71	35.08	743.63 (640.27 to 856.15)	3.9971	4.0056	-0.0084 (-0.14 to 0.12)	Dominated
Comprehension ability (baseline)	Normal	695.77	39.99	655.78 (572.95 to 746.36)	4.0083	4.0031	0.0053 (-0.29 to 0.36)	124,456
	Mild	677.12	37.10	640.02 (562.53 to 725.97)	4.3544	4.3597	-0.0053 (-0.10 to 0.07)	Dominated
	Moderate	804.44	39.68	764.76 (659.50 to 882.54)	4.1968	4.1342	0.0626 (-0.04 to 0.28)	12,207
Time post stroke (quartiles)	Severe	940.81	34.26	906.55 (635.42 to 1259.27)	3.6722	3.8511	-0.1789 (-1.06 to 0.32)	Dominated
	1 (shortest)	721.32	34.43	686.88 (603.04 to 781.92)	4.3279	4.3159	0.0120 (-0.32 to 0.36)	57,220
	2	763.36	44.19	719.17 (591.14 to 869.35)	4.6455	4.4419	0.2037 (-0.07 to 0.62)	3531
	3	752.31	36.27	716.04 (618.51 to 826.13)	4.0446	3.9685	0.0760 (-0.17 to 0.45)	9418
	4 (longest)	687.08	37.39	649.70 (544.46 to 772.47)	3.9071	3.8880	0.0191 (-0.12 to 0.18)	34,008

TABLE 43 Results of subgroup analyses comparing AC with UC: model-based analysis

Analysis	Subgroup	AC cost (£)	UC cost (£)	Incremental cost (£): AC vs. UC (95% CrI)	AC QALYs	UC QALYs	Incremental QALYs: AC vs. UC (95% CrI)	ICER (£ per QALY gained)
Word-finding difficulty (baseline)	Mild	40.17	0.00	40.17 (35.37 to 45.37)	4.2658	4.2564	0.0094 (-0.04 to 0.09)	4273
	Moderate	37.99	0.00	37.99 (31.95 to 44.72)	4.2874	4.3006	-0.0132 (-0.07 to 0.03)	Dominated
	Severe	35.08	0.00	35.08 (29.78 to 40.81)	4.0056	4.0137	-0.0081 (-0.14 to 0.10)	Dominated
Comprehension ability (baseline)	Normal	39.99	0.00	39.99 (33.87 to 46.81)	4.0031	4.0108	-0.0077 (-0.31 to 0.26)	Dominated
	Mild	37.10	0.00	37.10 (31.93 to 42.73)	4.3597	4.3595	0.0002 (-0.03 to 0.03)	185,500
	Moderate	39.68	0.00	39.68 (33.86 to 45.91)	4.1342	4.1360	-0.0018 (-0.08 to 0.08)	Dominated
	Severe	34.26	0.00	34.26 (23.83 to 46.67)	3.8511	3.7654	0.0857 (-0.19 to 0.45)	400
Time post stroke (quartiles)	1 (shortest)	34.43	0.00	34.43 (27.37 to 42.61)	4.3159	4.3211	-0.0052 (-0.08 to 0.07)	Dominated
	2	44.19	0.00	44.19 (37.28 to 51.70)	4.4419	4.5172	-0.0753 (-0.30 to 0.03)	Dominated
	3	36.27	0.00	36.27 (29.62 to 43.76)	3.9685	3.9979	-0.0294 (-0.41 to 0.33)	Dominated
	4 (longest)	37.39	0.00	37.39 (32.73 to 42.59)	3.8880	3.8632	0.0248 (-0.04 to 0.17)	1508

Appendix 16 Within-trial analysis: utility scores

TABLE 44 Mean and SD of accessible EQ-5D-5L scores by treatment group at each time point after missing EQ-5D-5L scores are imputed

Time point	EQ-5D-5L score, mean (SD)		
	UC (n = 101)	AC (n = 80)	CSLT (n = 97)
0 months	0.63 (0.23)	0.59 (0.26)	0.61 (0.24)
6 months	0.61 (0.24)	0.59 (0.26)	0.60 (0.28)
9 months	0.63 (0.23)	0.62 (0.23)	0.62 (0.25)
12 months	0.65 (0.23)	0.59 (0.24)	0.59 (0.25)

TABLE 45 Mean, SD and sample size of accessible EQ-5D-5L scores by treatment group at each time point for complete cases

Time point	EQ-5D-5L score, mean (SD); n		
	UC	AC	CSLT
0 months	0.63 (0.23); 101	0.59 (0.26); 80	0.61 (0.24); 96
6 months	0.61 (0.25); 89	0.59 (0.27); 71	0.60 (0.29); 84
9 months	0.63 (0.24); 86	0.63 (0.25); 64	0.62 (0.27); 80
12 months	0.66 (0.24); 87	0.60 (0.26); 62	0.59 (0.28); 77

TABLE 46 Mean and SD of accessible EQ-5D-5L scores by treatment group at each time point for participants who provided a complete response at every data collection point

Time point	EQ-5D-5L score, mean (SD)		
	UC (n = 81)	AC (n = 61)	CSLT (n = 73)
0 months	0.67 (0.19)	0.61 (0.25)	0.62 (0.25)
6 months	0.63 (0.25)	0.60 (0.27)	0.63 (0.27)
9 months	0.63 (0.25)	0.63 (0.25)	0.63 (0.27)
12 months	0.66 (0.24)	0.60 (0.26)	0.60 (0.28)

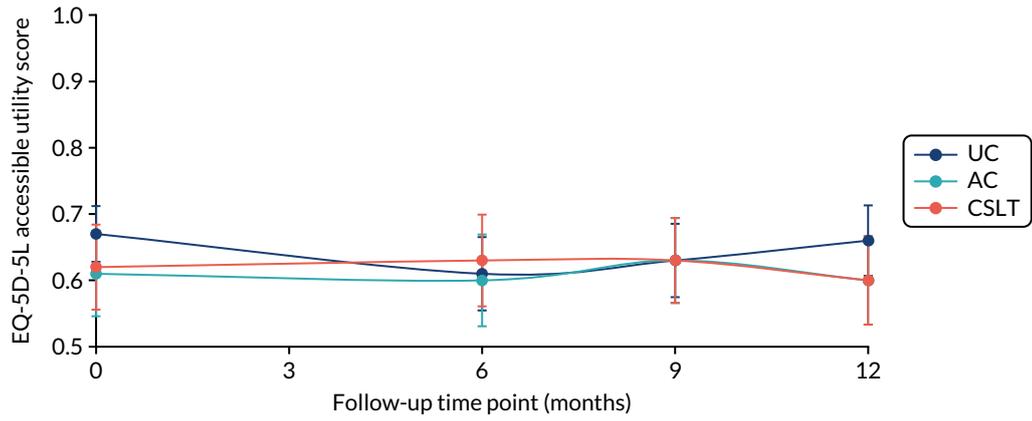


FIGURE 28 Means and 95% CIs of accessible EQ-5D-5L scores by treatment group at each time point for complete cases.

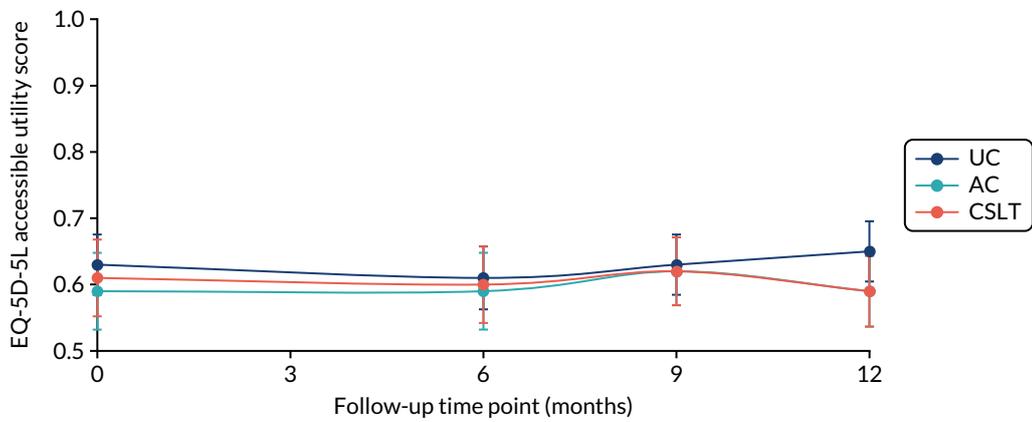


FIGURE 29 Means and 95% CIs of accessible EQ-5D-5L scores by treatment group at each time point for participants who provided a complete response at every data collection point.

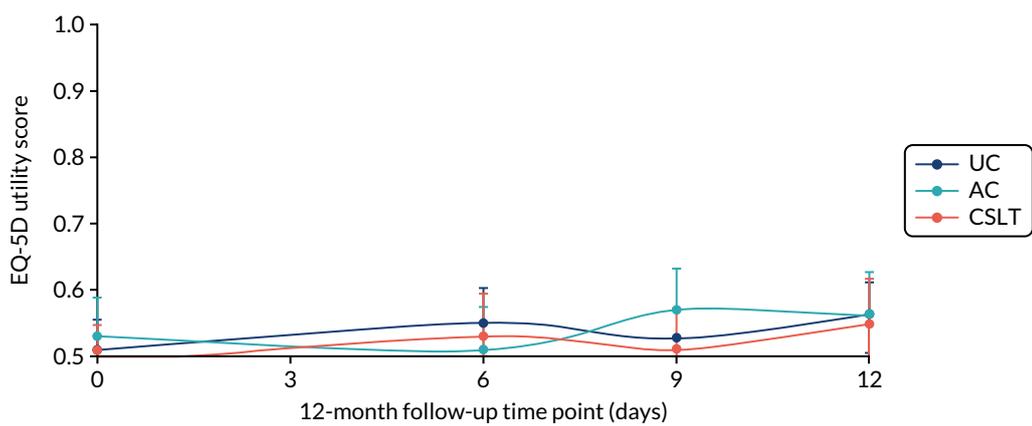


FIGURE 30 Means and 95% CIs of proxy EQ-5D-5L scores by treatment group at each time point (complete cases).

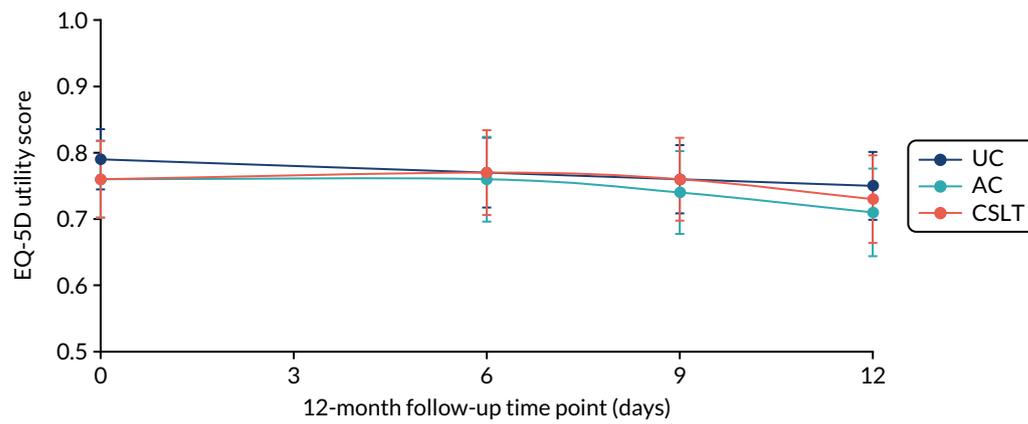


FIGURE 31 Means and 95% CIs of carer EQ-5D-5L scores by treatment group at each time point (complete cases).

Appendix 17 Within-trial analysis: histograms of cost distributions

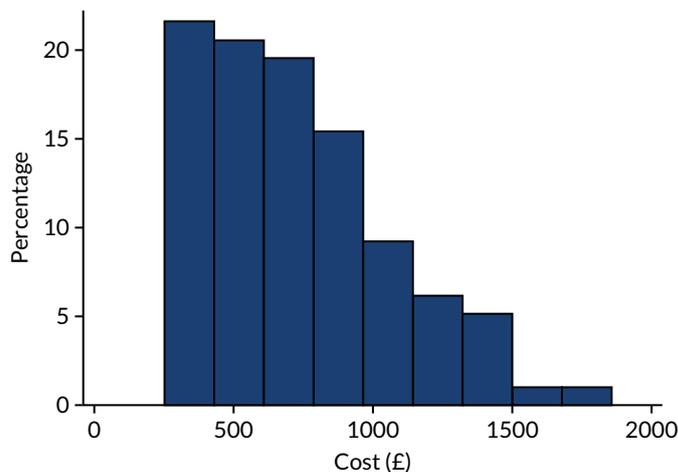


FIGURE 32 Histogram of CSLT costs.

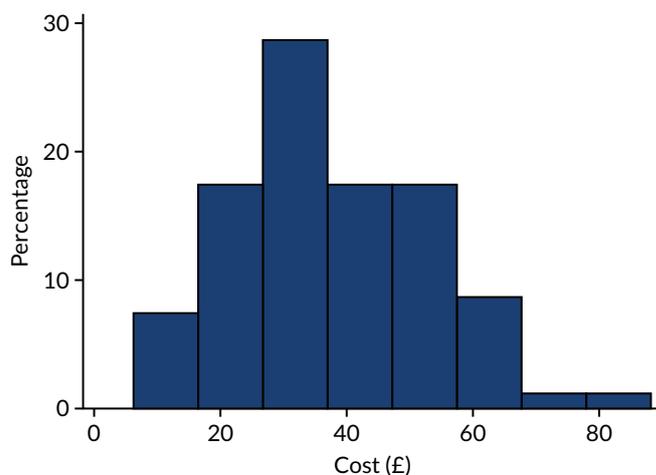


FIGURE 33 Histogram of AC costs.

Appendix 18 Within-trial analysis: base-case cost-effectiveness planes and cost-effectiveness acceptability curve

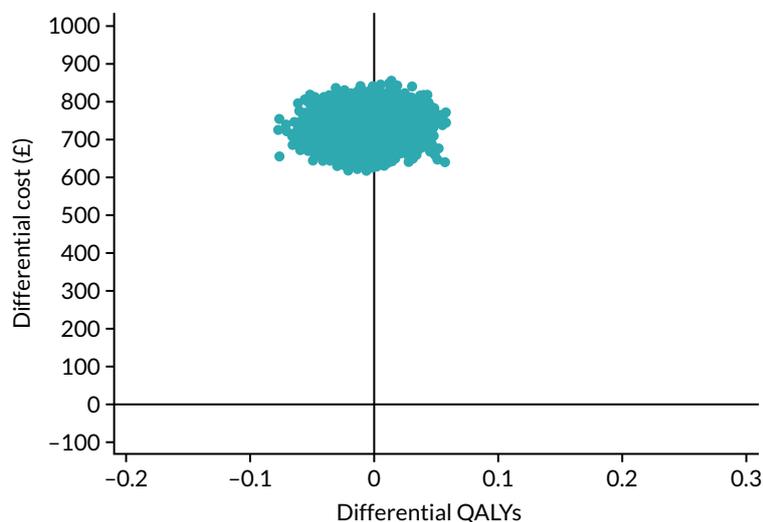


FIGURE 34 Cost-effectiveness plane: within-trial analysis, base case (CSLT vs. UC).

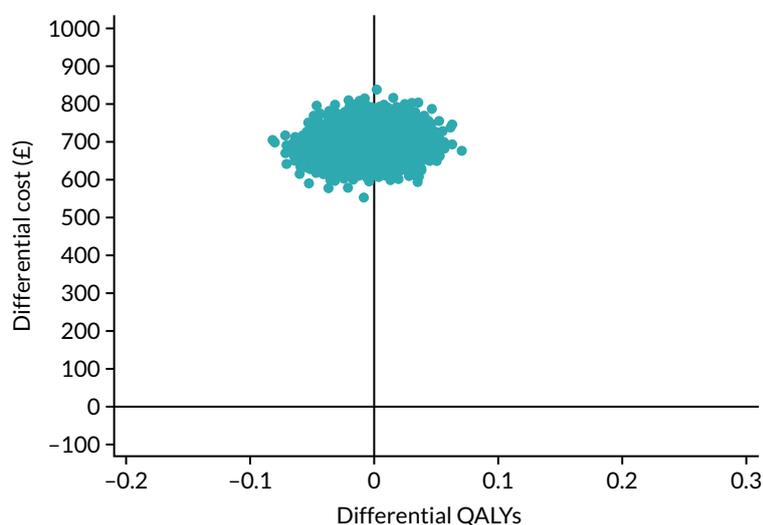


FIGURE 35 Cost-effectiveness plane: within-trial analysis, base case (CSLT vs. AC).

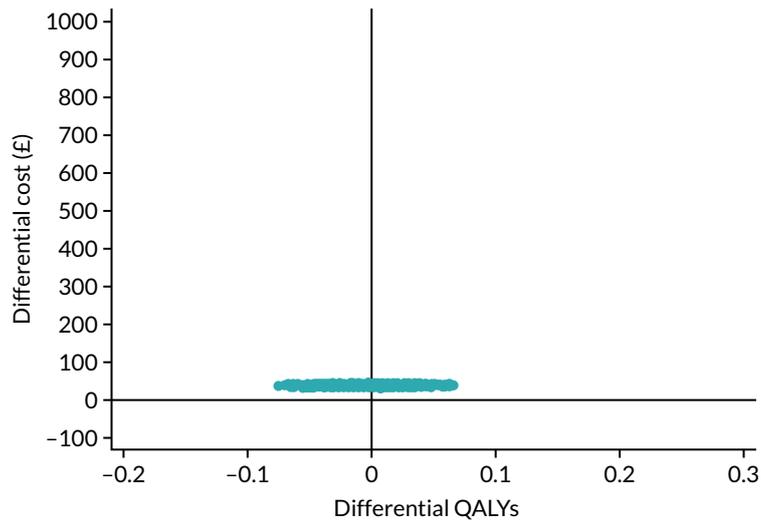


FIGURE 36 Cost-effectiveness plane: within-trial analysis, base case (AC vs. UC).

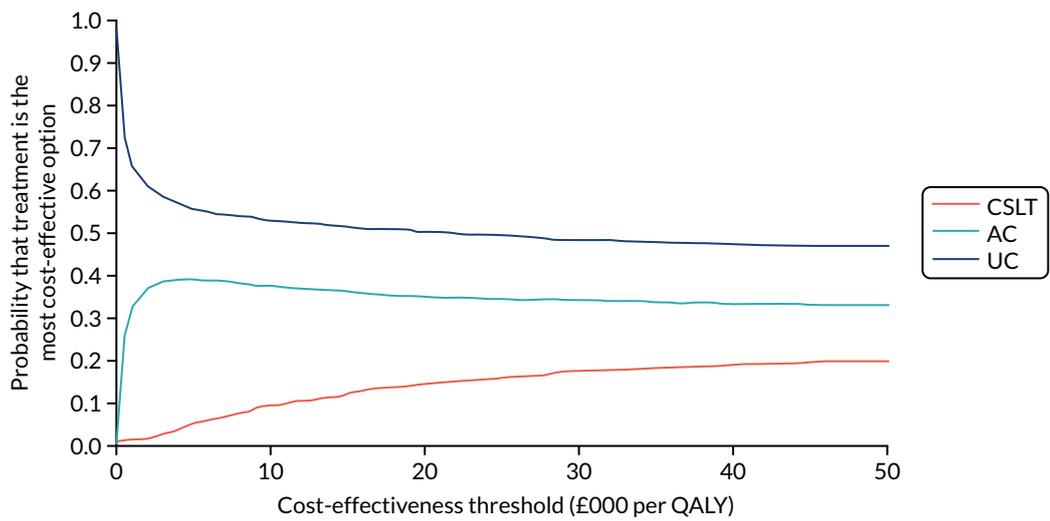


FIGURE 37 Cost-effectiveness acceptability curve: within-trial analysis, base case.

Appendix 19 Within-trial analysis: secondary analysis results

TABLE 47 Cost-effectiveness results from base-case and secondary within-trial analyses: CSLT vs. UC

Analysis	Incremental cost (£): CSLT vs. UC (95% CI)	Incremental QALYs: CSLT vs. UC (95% CI)	ICER (£ per QALY gained)
Base-case analysis (cross-walk)	732.37 (665.59 to 799.15)	-0.007 (-0.05 to 0.03)	Dominated
1. Complete-case analysis	732.37 (665.59 to 799.15)	0.014 (-0.03 to 0.06)	52,312
2. Using English EQ-5D-5L tariff	732.37 (665.59 to 799.15)	-0.007 (-0.05 to 0.03)	Dominated
3. Using carer proxy EQ-5D-5L measure	732.37 (665.59 to 799.15)	0.007 (-0.03 to 0.04)	104,624
4. Use EQ-5D-5L mapping of Hernandez-Alava <i>et al.</i> ⁶³	732.37 (665.59 to 799.15)	-0.011 (-0.05 to 0.03)	Dominated
5. Including carer QALYs	732.37 (665.59 to 799.15)	0.006 ^a	126,708
11. Broader perspective	789.60 (725.90 to 853.30)	-0.007 (-0.05 to 0.03)	Dominated

a Carer QALYs were weighted by a factor of 0.71, reflecting that 71% of patients had a carer at baseline, and carer QALYs were added to patient QALYs to calculate a total QALY gain (or loss). No CI can be derived from this method.

TABLE 48 Cost-effectiveness results from base-case and secondary within-trial analyses: CSLT vs. AC

Analysis	Incremental cost (£): CSLT vs. AC (95% CI)	Incremental QALYs: CSLT vs. AC (95% CI)	ICER (£ per QALY gained)
Base-case analysis (cross-walk)	694.65 (619.54 to 769.75)	-0.007 (-0.05 to 0.03)	Dominated
1. Complete-case analysis	694.65 (619.54 to 769.75)	0.006 (-0.04 to 0.05)	115,775
2. Using English EQ-5D-5L tariff	694.65 (619.54 to 769.75)	-0.013 (-0.05 to 0.02)	Dominated
3. Using carer proxy EQ-5D-5L measure	694.65 (619.54 to 769.75)	-0.003 (-0.04 to 0.03)	Dominated
4. Use EQ-5D-5L mapping of Hernandez-Alava <i>et al.</i> ⁶³	694.65 (619.54 to 769.75)	-0.019 (-0.06 to 0.02)	Dominated
5. Including carer QALYs	694.65 (619.54 to 769.75)	0.009 ^a	80,586
11. Broader perspective	751.80 (680.20 to 823.50)	-0.007 (-0.05 to 0.03)	Dominated

a Carer QALYs were weighted by a factor of 0.71, reflecting that 71% of patients had a carer at baseline, and carer QALYs were added to patient QALYs to calculate a total QALY gain (or loss). No CI can be derived from this method.

TABLE 49 Cost-effectiveness results from base-case and secondary within-trial analyses: AC vs. UC

Analysis	Incremental cost (£): AC vs. UC (95% CI)	Incremental QALYs: AC vs. UC (95% CI)	ICER (£ per QALY gained)
Base-case analysis (cross-walk)	37.72 (34.61 to 40.82)	-0.004 (-0.04 to 0.04)	Dominated
1. Complete-case analysis	37.72 (34.61 to 40.82)	0.004 (-0.04 to 0.05)	10,117
2. Using English EQ-5D-5L tariff	37.72 (34.61 to 40.82)	0.005 (-0.03 to 0.04)	7544
3. Using carer proxy EQ-5D-5L measure	37.72 (34.61 to 40.82)	0.012 (-0.02 to 0.05)	3143
4. Use EQ-5D-5L mapping of Hernandez-Alava <i>et al.</i> ⁶³	37.72 (34.61 to 40.82)	0.004 (-0.03 to 0.04)	9430
5. Including carer QALYs	37.72 (34.61 to 40.82)	-0.006 ^a	Dominated

a Carer QALYs were weighted by a factor of 0.71, reflecting that 71% of patients had a carer at baseline, and carer QALYs were added to patient QALYs to calculate a total QALY gain (or loss). No CI can be derived from this method.

Appendix 20 Within-trial analysis: subgroup analysis results

TABLE 50 Results of subgroup analyses comparing CSLT with UC: within-trial analysis

Analysis	Subgroup	Incremental cost (£): CSLT vs. UC (95% CI)	Incremental QALYs: CSLT vs. UC (95% CI)	ICER (£ per QALY gained)
Word-finding difficulty (baseline)	Mild	639.0 (555.9 to 722.1)	-0.014 (-0.07 to 0.04)	Dominated
	Moderate	832.3 (691.1 to 973.5)	-0.021 (-0.09 to 0.04)	Dominated
	Severe	776.9 (656.9 to 896.9)	0.027 (-0.05 to 0.11)	28,774
Comprehension ability (baseline)	Normal	685.5 (603.6 to 767.3)	-0.104 (-0.18 to -0.02)	Dominated
	Mild	671.9 (575.2 to 768.5)	0.021 (-0.03 to 0.07)	31,995
	Moderate	803.8 (662.9 to 944.7)	0.028 (-0.06 to 0.11)	28,707
	Severe	949.2 (601.2 to 1297.2)	-0.012 (-0.22 to 0.19)	Dominated
Time post stroke (quartiles)	1 (shortest)	716.5 (605.3 to 827.7)	-0.006 (-0.09 to 0.08)	Dominated
	2	760.2 (596.5 to 923.9)	-0.039 (-0.10 to 0.03)	Dominated
	3	759.2 (632.8 to 885.6)	0.026 (-0.06 to 0.11)	29,200
	4 (longest)	687.7 (559.9 to 815.5)	-0.013 (-0.07 to 0.04)	Dominated

TABLE 51 Results of subgroup analyses comparing CSLT with AC: within-trial analysis

Analysis	Subgroup	Incremental cost (£): CSLT vs. AC (95% CI)	Incremental QALYs: CSLT vs. AC (95% CI)	ICER (£ per QALY gained)
Word-finding difficulty (baseline)	Mild	599.2 (518.2 to 680.1)	-0.012 (-0.07 to 0.05)	Dominated
	Moderate	794.9 (608.7 to 980.9)	0.032 (-0.03 to 0.09)	24,715
	Severe	742.7 (608.7 to 875.7)	-0.015 (-0.08 to 0.05)	Dominated
Comprehension ability (baseline)	Normal	643.7 (542.0 to 745.5)	-0.036 (-0.14 to 0.06)	Dominated
	Mild	635.7 (522.2 to 749.3)	0.021 (-0.02 to 0.07)	30,080
	Moderate	764.5 (617.8 to 911.2)	0.006 (-0.07 to 0.08)	126,750
	Severe	916.9 (632.5 to 1201.4)	-0.164 (-0.31 to -0.02)	Dominated
Time post stroke (quartiles)	1 (shortest)	682.8 (560.2 to 805.3)	-0.001 (-0.09 to 0.09)	Dominated
	2	716.4 (526.0 to 906.8)	-0.076 (-0.14 to -0.01)	Dominated
	3	723.3 (559.0 to 887.6)	-0.018 (-0.10 to 0.07)	Dominated
	4 (longest)	650.6 (525.2 to 775.9)	0.052 (-0.02 to 0.12)	12,434

TABLE 52 Results of subgroup analyses comparing AC with UC: within-trial analysis

Analysis	Subgroup	Incremental cost (£): AC vs. UC (95% CI)	Incremental QALYs: AC vs. UC (95% CI)	ICER (£ per QALY gained)
Word-finding difficulty (baseline)	Mild	39.8 (34.4 to 45.2)	-0.008 (-0.06 to 0.05)	Dominated
	Moderate	37.5 (32.8 to 42.1)	-0.042 (-0.09 to 0.01)	Dominated
	Severe	34.2 (29.0 to 39.5)	0.041 (-0.05 to 0.13)	836
Comprehension ability (baseline)	Normal	41.7 (36.5 to 46.9)	-0.077 (-0.15 to 0.00)	Dominated
	Mild	36.2 (31.2 to 41.1)	0.001 (-0.05 to 0.05)	361,000
	Moderate	39.3 (34.0 to 44.6)	0.028 (-0.06 to 0.11)	1403
Time post stroke (quartiles)	Severe	32.2 (17.0 to 47.5)	0.106 (-0.10 to 0.32)	301
	1 (shortest)	33.7 (26.1 to 41.2)	0.013 (-0.08 to 0.11)	2592
	2	43.8 (37.3 to 50.3)	0.031 (-0.03 to 0.10)	1412
	3	35.9 (30.5 to 41.3)	0.021 (-0.08 to 0.12)	1709
	4 (longest)	37.1 (32.4 to 41.8)	-0.062 (-0.13 to 0.00)	Dominated

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

*This report presents independent research funded by the National Institute for Health Research (NIHR).
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the
Department of Health and Social Care*

Published by the NIHR Journals Library