University of Glasgow

Coleman, Rebecca (2026) *Exploring digital aspects of assessment and intervention in clinical neuropsychology.* D Clin Psy thesis

https://theses.gla.ac.uk/86328/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk



Exploring digital aspects of assessment and intervention in clinical neuropsychology

Rebecca Coleman, BA (Hons), MSc

Submitted in partial fulfilment of the requirements for the degree of

Doctorate in Clinical Psychology

School of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

September 2026

Table of Contents

List of Tables **5**

List of Figures **5**

Acknowledgements **6**

Chapter 1 Systematic Review **7**

Abstract **8**

Introduction **9**

Methods **13**

Search Strategy *15*

Screening Process *16*

Data Extraction Process *16*

Handling of Linked Reports *16*

Quality Appraisal *16*

Data Synthesis *17*

Data Presentation *17*

Results **18**

Search Results *18*

Study Characteristics *18*

Intervention Characteristics *22*

Effectiveness by Outcome Domain *22*

Methodological and Reporting Quality Appraisal *24*

Discussion **26**

Summary of Findings *26*

Relation to Existing Literature *26*

Strengths and Limitations *28*

Clinical and Research Implications *29*

Conclusion *30*

References **31**

Chapter 2 Major Research Project **37**

Plain Language Summary *38*

Abstract **39**

Introduction **40**

History of remote cognitive testing *40*

Acceptability of remote assessment 41
Feasibility, reliability, and validity of remote assessment 42
Current Study Aims 43

Method **45**
Study Design 45
Ethical Approval 45
Participants 45
Inclusion and Exclusion Criteria 46
Demographic Characteristics of Participants 46
Cognitive Measures 47
Research Procedure 50
Analysis Plan 50
Sample Size and Power 51

Results **52**
Preliminary Analyses 52
Variable Selection 52
Assessment of Normality 52
Descriptive Statistics for Primary Cognitive Variables 53
Primary Analysis: Consistency Between Administration Conditions 55
Intraclass Correlation Coefficients 55
Spearman's Rank-Order Correlations 57
Secondary Analysis: Differences by Administration Condition, Age Group, and Testing Order 59
Effect of Administration Condition 59
Effects of Age Group and Testing Order 59
Age Group by Testing Order Interaction 60
Effect of Education 61
Summary 61

Discussion **63**
Summary of Results 63
Relation to Existing Findings from the Literature 63
Strengths and Limitations of the Study 66
Theoretical Implications 68

References **70**

Appendices **75**
Appendix A. PRISMA reporting guidelines 75
Appendix B. Search Strategies 78
Appendix C. CCAT Quality Appraisal Ratings for all six included RCTs: Lead and Second Reviewer 81

Appendix D: STROBE checklist 82
Appendix E: Final approved MRP proposal 84
Appendix F: Ethics Approval Letter 85
Appendix G: Participant Information sheet 86
Appendix H: Consent form 87
Appendix I: Data Analysis Plan 88
https://osf.io/7deja/files/puxrm 88
Appendix J: Data Analysis Process 89
Appendix K: Data Management Plan (DMP) 90
*Appendix L: G*Power Sensitivity Analysis for Correlations and Effect Size (Planned Sample, N = 20) and Achieved Sample (N = 16)* 91
Appendix M: Practice Effects Descriptives 93
Appendix N: Condition x Age Group Interaction: Full Univariate Results 94
Appendix O: Sample MRP Reflexivity Journal Entry 95

List of Tables

Table 1 Study Eligibility Criteria 13
Table 2 Characteristics and key results of each included study 20
Table 3 Demographic characteristics of participants 47
Table 4 Cognitive Measures 48
Table 5 Median (IQR range) for primary cognitive variables by administration condition and age group 53
Table 6 Intraclass correlation coefficients (single measures) between in-person and remote administration 55
Table 7 Spearman's rank-order correlations between in-person and remote scores 58
Table 8 Estimated marginal means (controlling for Education Years) for Age Group and testing order, for measures showing a significant or near significant interaction 61

List of Figures

Figure 1 PRISMA 2020 flow diagram of study identification and selection (PRISMA template; Page et al., 2021), reproduced below illustrating the review's figures. 19

Acknowledgements

First, I would like to thank my research supervisor, Dr Jessica Fish, whose input and support has been invaluable throughout this project. Your feedback was always constructive and encouraging, your patience was constant, and your guidance has taught me as much about doing research well as it has about this specific project. I could not have asked for a more invested supervisor, and I am genuinely grateful for the time and care you gave this project.

I would also like to thank my field supervisor, Dr Ruth Sumpter, for grounding this project in clinical practice throughout, for the insights from the field that helped keep the research relevant, and for your patience with the practical realities of recruitment and data collection, and the reassurance you offered whenever things felt difficult.

This project has, at times, been a genuine challenge to balance alongside everything else the doctorate demands, but it has equally been an enjoyable part of my training, and I have learned a great deal in the process.

To my partner and my family, thank you for your steady support over the past three years. Between us, we have racked up more miles on the A9 than I can calculate, whether it was me heading home or them making the trip up to see me. A particular thank you to my parents, who have made that journey more times than I can count, and who have been a constant source of support throughout not just this thesis, but my entire education.

Chapter 1 Systematic Review

The effectiveness, feasibility, and acceptability of remotely delivered cognitive behavioural therapy (CBT) for adults with acquired brain injury: A systematic review

Prepared in accordance with the author requirements for Archives of Clinical Neuropsychology; https://academic.oup.com/acn/pages/author-guidelines

Key words: acquired brain injury; traumatic brain injury; stroke; cognitive behavioural therapy; telehealth; remote delivery

Abstract

Background: Acquired brain injury (ABI) is commonly associated with persistent psychological difficulties, including depression, anxiety, and sleep disturbance, which can further impact recovery and quality of life. Cognitive Behavioural Therapy (CBT) is an effective, first-line treatment for these difficulties and remote delivery formats can improve access for people with ABI who experience mobility, fatigue, and geographical barriers to accessing face-to-face appointments. However, no review has synthesised evidence specifically exploring the effectiveness, feasibility and acceptability of remotely delivered CBT for adults with ABI.

Methods: A systematic review was conducted following PRISMA 2020 guidelines. MEDLINE, Embase, CENTRAL and PsycINFO were searched until 29/06/2026 and relevant data extracted. Eligible studies consisted of randomised controlled trials (RCTs), including pilot and feasibility RCTs, evaluating CBT or CBT-derived interventions that are clearly defined and delivered remotely to adults with ABI. Quality appraisal was conducted using the Crowe Critical Appraisal Tool (CCAT). Findings were synthesised and grouped by outcome domain.

Results: Six RCTs met inclusion criteria, with individual study sample sizes ranging from 24-126 participants (total n=349). The RCTs measured a variety of outcomes including insomnia, depression, fatigue, functional cognitive symptoms, and post-concussion avoidance behaviour. Remotely delivered CBT was most consistently effective for sleep and fatigue, particularly post-treatment, whereas evidence for depression, anxiety and memory-related difficulties was more mixed. Feasibility and acceptability were high across studies, with strong recruitment, retention, treatment credibility and session adherence reported. Interventions were delivered via videoconferencing, blended online / face-to-face modalities, and telephone formats. CCAT methodological and reporting quality appraisal scores were generally high (82.5%-97.5%), though some studies gave limited detail on ethics and sampling, especially feasibility trials.

Discussion: In conclusion, remotely delivered CBT appears feasible and acceptable for adults with ABI, and effective for improving sleep, though evidence for fatigue and mood outcomes is more mixed. Larger trials with longer follow-up are needed to confirm the effectiveness of remote CBT for ABI and identify who benefits most from which delivery format. This review was unfunded. Prospero registration number: CRD 420261447368.

Introduction

Acquired brain injury (ABI) refers to brain injuries sustained after birth, resulting from a wide range of causes, including trauma, stroke, infection, hypoxia, or tumour for example; and is one of the leading causes of long-term disability worldwide (Feigin et al., 2021). In addition to the physical and cognitive impact associated with ABI, individuals also frequently experience significant psychological difficulties, including depression, anxiety, and disrupted sleep, which can persist after the initial injury, subsequently affecting overall quality of life, social functioning, relationships and recovery (Mar et al., 2011). More specifically, sleep disturbance has a high prevalence within this population, affecting up to 50% of people with TBI (Mathias & Alvaro, 2012) and 38% of stroke survivors (Baylan et al., 2020), and can also exacerbate comorbid anxiety, depression and fatigue (Ouellet et al., 2015).

Despite the high prevalence of psychological difficulties following ABI, access to targeted treatment may be limited. Sleep disturbance provides a clear illustration of this treatment gap. For example, in a survey of 452 people with TBI, approximately half reported insomnia symptoms, however, almost 60% of those meeting the criteria for insomnia were not receiving treatment for their sleep difficulties (Ouellet et al., 2006). More recently, Ford et al. (2023) noted that up to one third of people with stroke or TBI may meet diagnostic criteria for insomnia disorder and identified limited availability of CBT-I as a key barrier to wider implementation. This suggests a clinically meaningful treatment gap and highlights the need for accessible, evidence-informed psychological interventions following ABI.

Research has highlighted that, across people with head injuries of all severities, in the chronic phase disability is more strongly linked to psychological difficulties than to physical or injury severity factors (Whitnall et al., 2006). Further, Bulloch and colleagues (2015) similarly found increased depressive symptoms measured using the PHQ-9 across a broad range of neurological populations, drawing on a national Canadian survey of people living with neurological conditions. Within TBI specifically, Scholten et al. (2016) whose review included studies across all severities of TBI, found that anxiety and depression become more common over time; anxiety disorders roughly increased from 21% in the first year to 36% long term, while depressive disorders doubled from 17% to 43%. Similarly, stroke survivors often experience these difficulties also, with around one in four experiencing anxiety within

the first five years, and one in three experiencing depression within the first year (Stroke Association, n.d.).

Cognitive Behavioural Therapy (CBT) is a first-line, evidence based psychological treatment for depression, anxiety and sleep difficulties in the general population and is increasingly being applied for people with ABI (Zhu et al., 2026), though evidence on adaptations of CBT protocols specifically for this population remains limited. Additionally, access to face-to-face CBT is limited by long waiting lists and geographical barriers and these barriers are exacerbated for people with ABI due to fatigue, mobility limitations (Burgess et al., 2024), and transport difficulties (Avramovic et al., 2023). Remote or digitally delivered CBT, including videoconferencing and telephone-based formats offers a means of improving access to psychological therapy without these practical barriers.

Evidence for remote CBT is growing steadily. Andersson and colleagues (2019) found that internet delivered CBT (iCBT) performs comparably to face-to-face CBT. Similarly, Mehta et al. (2019) conducted a meta-analysis of 25 RCTs of iCBT across a range of chronic health conditions (e.g., fibromyalgia, chronic pain and spinal cord injury) and found that therapist guided iCBT led to better depression outcomes than self-guided iCBT, though both approaches were similarly effective for anxiety. A meta-analysis of CBT for depression and anxiety across eleven neurological conditions found smaller and less reliable effects for TBI than for conditions such as MS or Parkinson's, a finding based on relatively few TBI trials (Gandy et al., 2024). This suggests that the evidence base for CBT following TBI may be less developed in comparison to other neurological conditions.

Further, clinician involvement appears to meaningfully improve engagement with remote CBT. A recent meta-analysis looking at psychiatric disorders found completion rates of 79% for therapist guided remote CBT compared with 48% for self-guided formats (Bakanaitė et al., 2025). Although not specific to ABI, this may be because facilitated sessions allow for personalisation with regards to pacing and content, supporting therapeutic alliance in ways that self-directed programmes cannot easily replicate. This may be particularly relevant for people with ABI for whom cognitive difficulties can make independent engagement with self-led material harder to sustain. Moreover, a systematic review of digital health interventions for TBI shows overall positive, feasible outcomes, though most existing interventions used a single delivery modality rather than multi-modal approaches. (Avramovic et al., 2023). Because the evidence base for telephone delivered interventions is

already comparatively well established, the present review excludes interventions delivered solely by telephone with no video or online component, in order to focus on video and internet-based delivery formats where the evidence remains less developed and which is arguably of strongest relevance to contemporary practice.

Several authors have raised concerns about the accessibility of remote and internet-based interventions for people with ABI, given the cognitive impairments associated. Miao et al. (2022) found that remote ABI interventions commonly excluded participants with cognitive, sensory or physical comorbidities, highlighting a tension between the accessibility remote delivery is intended to offer and the realities of living with ABI. However, Bergquist et al. (2008) found in a pilot feasibility study that people with moderate to severe TBI and associated memory impairment were able to learn to use an internet based cognitive rehabilitation programme independently, suggesting that appropriate support can help overcome accessibility barriers.

Seery and colleagues (2026) recently updated an earlier meta-analysis (Stalder- Lüthy et al. 2013) and synthesised evidence for psychological interventions for depression more than one year following an ABI. The review included interventions delivered in individual, group and combined individual/group formats, as well as remotely via internet, telephone and telehealth-based approaches, across a range of psychological therapies including CBT, mindfulness-based interventions, acceptance and commitment therapy (ACT), behavioural activation and psychoeducation. Their review found an overall small, significant effect (SMD = -0.31), which became non-significant after adjusting for dependent effect sizes and small study bias; TBI only subgroup analyses were non-significant, while longer treatment duration (>20 sessions) and individual (rather than group) delivery were associated with stronger effects. Unlike Seery et al.'s (2026) review, which included but was not restricted to remote delivery and considered any psychological intervention for depression, the present review focuses specifically on CBT and clearly CBT-derived interventions delivered remotely, restricted to randomised controlled trials (RCTs), and also considers a broader range of outcomes, including sleep, feasibility, and acceptability, with no restriction on minimum time since injury.

Rationale and Aims

CBT is an effective, recommended treatment for depression, anxiety and sleep difficulties, and remote delivery appears broadly comparable to face-to-face formats where directly

compared. However, the evidence base for CBT specifically within ABI populations remains fragmented, based largely on single trials and reviews that do not synthesise effectiveness, feasibility, and acceptability of remotely delivered CBT as a distinct focus. The present review therefore is more narrowly focused on this, restricted to CBT and clearly CBT-derived interventions for ABI, aiming to identify, appraise and synthesise the evidence on the effectiveness, feasibility, and acceptability of remotely delivered CBT for adults with ABI, in order to inform the development and provision of accessible psychological interventions for this population.

Review Questions

1. What is the effectiveness of remotely delivered CBT, relative to comparator conditions, in improving psychological outcomes (e.g., anxiety, depression, sleep) in adults with ABI?
2. How feasible and acceptable is remotely delivered, facilitator-led CBT for adults with ABI (e.g., recruitment, adherence, retention, and participation satisfaction)?

Methods

The protocol for this review was registered on PROSPERO (ID: CRD 420261447368). The review follows the PRISMA reporting guidelines (Page et al., 2021); the completed checklist can be found in Appendix A. Note that three amendments were made to the protocol following registration, agreed within the review team:

1. The intervention could be delivered by any healthcare professional, or individual trained to deliver the intervention (including under supervision) rather than a psychologist or psychological therapist specifically
2. No minimum proportion of remote delivery was required, replacing the initial ≥50% threshold, due to difficulty ascertaining the exact proportion of remote delivery reported in many trials.
3. The Crowe Critical Appraisal Tool (CCAT) was used in place of the initially specified Cochrane Risk of Bias 2 tool (RoB-2). Although RoB 2 is the established tool for assessing risk of bias in randomised trials, the present review included both full scale and pilot / feasibility RCTs, for which the breadth of reporting and methodological detail varied. The CCAT was therefore selected to provide a structured appraisal of methodological and reporting quality across all included studies, according to the features and purpose of each study design, rather than to generate a stand-alone risk of bias judgement. The CCAT has demonstrated construct validity and good inter-rater score reliability in prior evaluations (Crowe et al., 2011; Crowe et al., 2012). RoB 2 was also considered less feasible within the available review resources, as it requires substantial training and calibration, and published evaluations have reported low initial inter-rater agreement, although review specific implementation guidance may improve consistency (Minozzi et al., 2020, 2022).

Eligibility Criteria

Table 1 *Study Eligibility Criteria*

Studies were included in this review if they met the criteria in Table 1:

Inclusion criteria

Population	Adults (≥18 years) with acquired brain injury (ABI), including injuries associated with traumatic brain injury (TBI), stroke or other cause.

Intervention	CBT, or a clearly CBT-derived approach (e.g. cognitive therapy, cognitive restructuring, problem-solving therapy); can be delivered by any healthcare professional, or another individual trained to deliver the intervention (including under supervision), in either individual or group format. Remote delivery (e.g., videoconferencing) must be a primary or clearly identifiable component of the intervention.
Comparator	Any comparator / control group e.g., treatment as usual (TAU groups may differ between studies), standard rehabilitation care, psychoeducation, wait list control, social / support group, in-person / face-to-face CBT.
Outcomes	Feasibility / Acceptability / Effectiveness (recruitment, adherence, retention / attrition, satisfaction); anxiety (e.g., GAD-7); depression (e.g., PHQ-9); sleep (e.g., ISI) at any reported time point.
Study Design	Randomised controlled trials (RCTs), including pilot / feasibility RCTs.
Setting / Context	Any country or healthcare setting. Published, English language studies.

Exclusion criteria

Population	Children / adolescents (under 18 years); other neurological or neurodegenerative conditions (e.g. dementia, multiple sclerosis, Parkinson's); studies where ABI participants cannot be separated from a mixed clinical sample; studies focused solely on caregivers' / family members rather than individuals with ABI.
Intervention	Fully self-guided digital interventions with no facilitator input; music or art therapy; interventions delivered solely by telephone with no video / online component, interventions that clearly draw on a theoretical model distinct from CBT (e.g., Cognitive Analytical Therapy, Counselling, Psychodynamic, Systemic/family therapy, EMDR. Schema Therapy); interventions directed solely towards cognitive function (i.e., attention/memory/executive function strategies) or

	compensatory aids (i.e., memory aids, calendars, reminders); physical motor rehabilitation; medical/pharmacological treatment; virtual reality-based interventions.
Comparator	—
Outcomes	—
Study Design	Case studies / series; qualitative only studies; study protocols without results; non-randomised studies including single-arm feasibility / acceptability / pilot studies.
Setting / Context	—

The current review was restricted to RCTs (including pilot and feasibility RCTs) because randomisation controls for confounding and selection bias, providing the strongest available evidence for causal claims about treatment effectiveness, an important consideration given the heterogeneity of presentation, recovery trajectory and comorbidity within ABI populations.

Search Strategy

A literature search was conducted from database inception to 29/06/2026 using MEDLINE, Embase, CENTRAL (Cochrane Central Register of Controlled Trials), and PsycINFO, with no date restrictions and an English language restriction applied.

A specialist university librarian was consulted on the development of the search strategy, including the review of the final search terms and syntax. A combination of search terms related to acquired brain injury (e.g., "head trauma" OR "stroke" OR "hypoxic brain" OR encephalitis), as well as terms relating to Cognitive Behavioural Therapy (CBT) (e.g., "Cognitive Restructuring" OR "Cognitive Techniques"), and remote (e.g., "telehealth" OR "internet-based" OR "online" OR "videoconference"), were used across databases. Truncation was used (e.g., brain injur*, teletherap*, cognitive therap*) to maximise sensitivity and capture all relevant research. Each database was searched according to its own subject heading thesaurus; full search strategies and line by line syntax per database are reported in Appendix B.

Screening Process

All records were imported into Rayyan (Ouzzani et al., 2016) and duplicates removed automatically. The primary researcher then reviewed the titles and abstracts and included or excluded based on the eligible criteria. The first 70 records were screened jointly by two reviewers with 100% agreement. 20 full texts of all potentially eligible studies were then independently screened by two reviewers from the review team, with any disagreement resolved through discussion. A third reviewer also independently completed quality appraisal for all six final included studies (see Quality Appraisal below).

Data Extraction Process

The primary researcher completed data extraction of six included RCTs, recording: title; author(s) and year of publication; DOI; country; study design and setting; population characteristics; sample size; intervention description; primary outcome(s); secondary outcome(s); type of analysis; main results; effect size; follow-up time-points; feasibility/acceptability measures and ethical considerations.

Handling of Linked Reports

Where studies reported on overlapping trial data across multiple publications (e.g., the Ymer et al. CBT-SF trial, which incorporated data from an earlier underpowered internal pilot study, Ymer et al., 2021), the most complete and recent report was used as the primary source for data extraction and quality appraisal.

A matching DOI, then journal / volume / page range, then trial registration number was used to identify the primary index report. Earlier or linked reports were referred to only to supplement missing methodological detail and were not double counted as separate studies in the synthesis.

Quality Appraisal

Methodological and reporting quality of the included studies was appraised using the Crowe Critical Appraisal Tool (CCAT; Crowe, 2013). The CCAT was selected because the review included both full scale RCTs and pilot / feasibility RCTs. Although all included studies were randomised, pilot and feasibility trials are designed primarily to assess whether a larger trial is practical and acceptable, rather than to provide definitive evidence of treatment effectiveness. They may therefore report less detail on features typically expected in full scale efficacy trials, such as allocation concealment, sample size justification, and clinical outcome analyses. A broad critical appraisal tool was therefore considered appropriate for

applying a consistent framework across the included studies whilst taking account of the purpose and design of each study.

The CCAT is a general critical appraisal tool that evaluates methodological and reporting quality across eight domains: preliminaries, introduction, design, sampling, data collection, ethical matters, results, and discussion. Each domain is rated from 0 (poor) to 5 (excellent) giving a total score out of 40, which was converted to a percentage for descriptive presentation. The tool prompts consideration of potential sources of bias, confounding, and imprecision within relevant domains. However, CCAT total scores were interpreted as an overall quality appraisal and were not used as formal risk of bias judgements.

The CCAT has shown construct validity (Crowe & Sheppard, 2011) and has demonstrated good inter-rater score reliability (Crowe et al., 2011). A second reviewer independently quality appraised all six included RCTs using the CCAT. Ratings were compared at domain and total score level and were generally consistent, within 1-2 points for five of the six studies, with a maximum discrepancy of 4 points (10%) for Fisher et al. (2021), discrepancies were resolved through discussion. As only six studies were included, reviewer consistency is reported descriptively rather than using a formal reliability statistic.

Data Synthesis

A narrative synthesis was conducted; a meta-analysis was not planned given the anticipated heterogeneity between studies in intervention format, comparator type and outcome measurement. Studies were grouped by outcome domain (i.e., feasibility/acceptability, anxiety, depression, sleep) and findings were compared within each domain for consistency of direction and magnitude of effect, whilst taking into account differences in comparator type, study design, and methodological and reporting quality as appraised using the CCAT. Where multiple linked reports existed for the same trial, findings were synthesised as a single dataset to avoid double counting.

Data Presentation

Table 2 summarises the characteristics and key results of each included study, including CCAT methodological and reporting quality appraisal scores (lead reviewer). Full domain ratings and reviewer score comparisons are presented in Appendix C.

Results

Search Results

The search identified 1,454 records across MEDLINE, Embase, CENTRAL and PsycINFO. After removing 487 duplicates, 967 records were screened by title and abstract, of which 889 were excluded as not eligible. 78 full text articles were screened for eligibility, of which 72 were excluded (wrong intervention, n = 24; wrong study design, n = 14; delivery not remote/video, n = 12; wrong publication type, n = 8; registered but not yet completed, n = 7; conference abstract only, n = 3; wrong population, n = 2; duplicate/preliminary report, n =1; secondary analysis of parent RCT, n = 1); six RCTs met inclusion criteria and were retained for narrative synthesis (Figure 1).

Study Characteristics

The six included RCTs were published between 2021 and 2025 and conducted in the Netherlands (k = 1), United States (k = 2), Canada (k = 2), and Australia (k = 1). Sample sizes ranged from 24 (Rioux et al., 2024) to 126 participants (Ymer et al., 2025), with a total of 349 participants across all studies. Four trials included mixed acquired brain injury samples (stroke and / or TBI) (Ford et al., 2023; Fisher et al., 2021; Raina et al., 2022; Ymer et al., 2025), while two focused on persistent symptoms following mild TBI (mTBI) or concussion (Rioux et al., 2024; Silverberg et al., 2022). Study designs ranged from full scale, multicentre RCTs (Ford et al., 2023; Ymer et al., 2025) to smaller pilot and feasibility trials (Fisher et al., 2021; Raina et al., 2022; Rioux et al., 2024; Silverberg et al., 2022). Full study characteristics are summarised in Table 2.

Figure 1 PRISMA 2020 flow diagram of study identification and selection (PRISMA template; Page et al., 2021), reproduced below illustrating the review's figures.

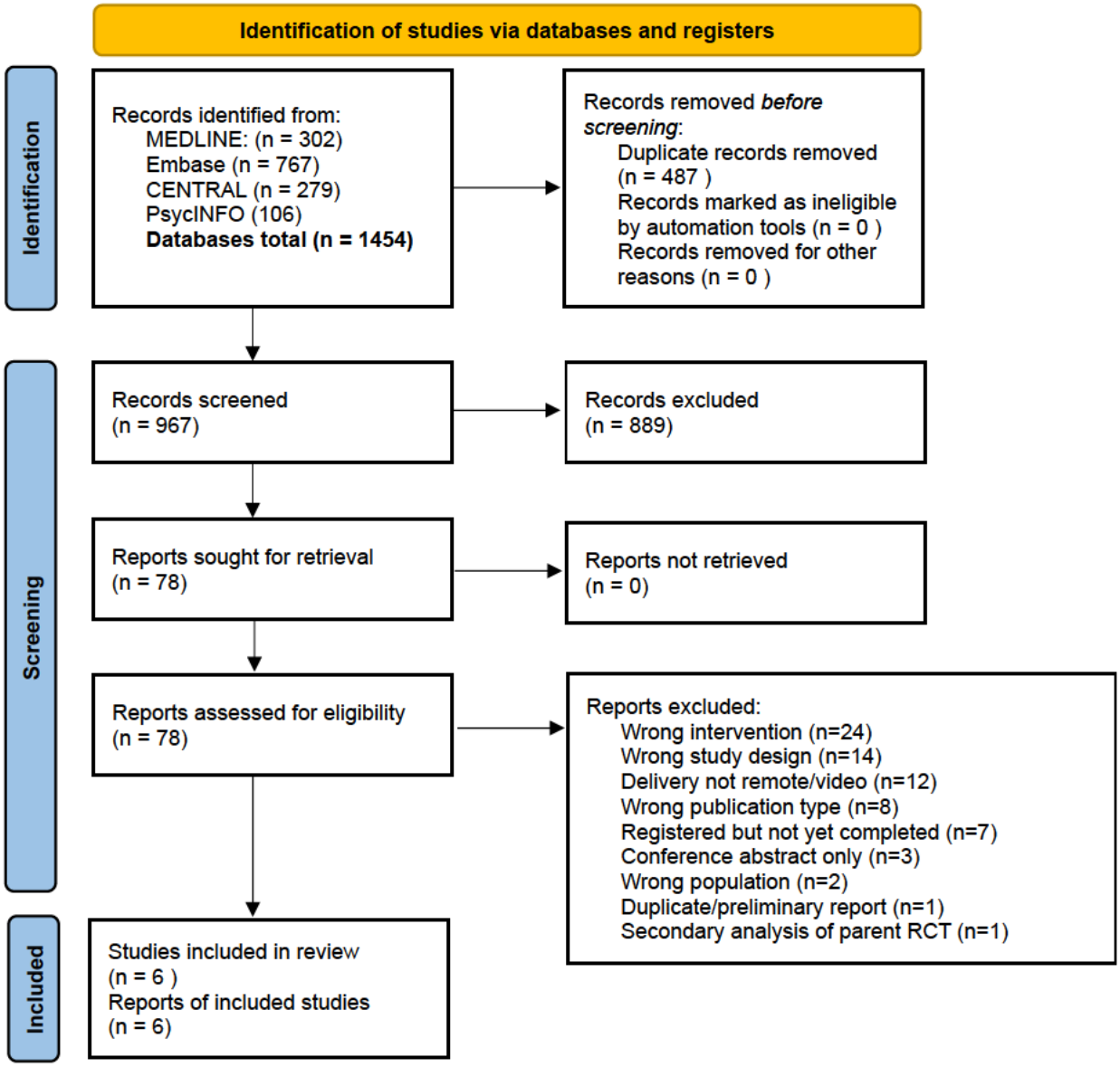


Table 2 *Characteristics and key results of each included study*

Study (Country)	Design; N	Population	Intervention (deliverer; remote modality; dose)	Comparator	Primary outcome(s)	Key findings	CCAT score (%)
Ford et al., 2023 (Netherlands)	Multicentre RCT; N=52 total (27 eCBT-I vs 25 TAU)	Adults with insomnia and stroke or TBI	Blended eCBT-I (6 weekly guided online sessions + 2 face-to-face) with a 6-week follow-up; healthcare psychologist; Mind district eHealth platform	Treatment as usual	Insomnia Severity Index (ISI)	eCBT-I > TAU on ISI post-treatment (d=0.96) and follow-up (d=0.78); greater clinical remission; effective in stroke but not TBI subgroup; no significant benefits for fatigue/mood/cognition	90.0
Fisher et al., 2021 (USA)	Pilot RCT (waitlist-controlled, ongoing); N=33 consented (n=18 remote-completer subsample analysed)	Moderate-severe TBI with depressive symptoms	CBT-TBI (12-week individual, manualised); master's/doctoral facilitators; zoom videoconferencing / telephone	12-week waitlist	Feasibility, acceptability and safety of remote delivery	High attendance/retention (93% remote period); high satisfaction (STTS-R therapy subscale, M=27.1/30; 64% 'very satisfied' 29% 'satisfied'); safety monitoring successfully implemented	82.5
Raina et al., 2022 (USA)	Pilot RCT; N=41 (20 MAX vs 21 HE)	Adults with TBI and clinically significant fatigue	MAX (problem-solving therapy + energy conservation); occupational therapist; 30min 2x	Health education (HE) i.e., diet, exercise, energy conservation	Modified Fatigue Impact Scale (mFIS)	MAX > HE on fatigue impact (d=0.87), fatigue experience (d=0.75) and fatigue severity (d=0.61) at 8-week follow-up; no significant between group difference in objective or subjective participation	95.0

			per week 8 weeks, live online sessions				
Rioux et al., 2024 (Canada)	Pilot feasibility RCT; N=24 (11 CBT vs 13 Cognitive Rehab)	Functional cognitive disorder after concussion	Manualised individual CBT (cognitive reappraisal, exposure); graduate clinical psychology trainees; Zoom video platform; 11 weekly sessions	Cognitive rehabilitation	Feasibility; Multifactorial Memory Questionnaire-Satisfaction (MMQ-S)	Feasibility benchmarks mostly met; both arms improved similarly on MMQ-S (d=-0.05); qualitative data favoured CBT-specific elements	95.0
Silverberg et al., 2022 (Canada)	Randomised feasibility trial; N=73 (37 GET vs 36 Pacing	Persistent post-concussion symptoms with high avoidance /endurance coping	Manualised; Graded exposure therapy (GET); psychology graduate students / postdoctoral + occupational therapist dyad; On Call Health videoconferencing; 1-hour sessions with each provider 8 weeks = total 16 therapy hours	Pacing + (active comparator)	Feasibility (accrual, credibility, fidelity, adherence, retention); Fear Avoidance Behaviour (FAB-TBI)	Most feasibility targets met; GET > Pacing+ for avoidance behaviour (d=0.81); Pacing+ non-significant advantage for endurance (d=0.39)	97.5
Ymer et al., 2025 (Australia)	RCT, 2:1 allocation; N=126 (86 CBT-SF vs 40 HE)	Adults with stroke / TBI and clinically significant sleep disturbance/fatigue	8-week CBT-SF (7 modules; 8 x 1-hour individual sessions); delivered by clinical neuropsychologists; face-to-face or videoconferencing (76%)	Health education (HE)	Pittsburgh Sleep Quality Index (PSQI)	CBT-SF > HE for PSQI (β= -1.50, $p<.001$) and ISI (d=0.95) post-treatment; between group differences were not maintained at follow-up; earlier within group improvements in fatigue and mood	97.5

Note. CCAT scores represent an appraisal of methodological and reporting quality across eight domains; they are not formal risk of bias ratings.

Intervention Characteristics

Interventions varied in CBT approach, target symptoms and mode of delivery, although all were facilitator-led and delivered individually. Three trials targeted sleep and / or fatigue; blended CBT-I (Ford et al., 2023), a problem-solving based fatigue self-management intervention (MAX; Raina et al., 2022), and a manualised CBT for sleep and fatigue (CBT-SF; Ymer et al., 2025). Two trials targeted post-concussion symptoms: cognitive reappraisal and exposure-based CBT for functional cognitive disorder (Rioux et al., 2024), and graded exposure therapy targeting avoidance behaviour, compared with a pacing-based intervention (Silverberg et al., 2022). One trial evaluated CBT for depression adapted for TBI (CBT-TBI; Fisher et al., 2021), reporting on the feasibility of its transition to remote delivery during the COVID-19 pandemic. Interventions were delivered by clinical psychologists, neuropsychologists, occupational therapists, or supervised psychology trainees / graduate students, via videoconferencing, blended eHealth platforms, or a combination of videoconferencing and telephone. Session number ranged from six to twelve individual sessions delivered over 6-12 weeks.

Effectiveness by Outcome Domain

Sleep

Two trials targeting sleep found that remotely delivered CBT produced significantly greater improvements than comparators immediately post-treatment. Ford et al. (2023) reported a significantly larger reduction in insomnia severity for blended eCBT-I compared with treatment as usual (d=0.96), maintained at 6 week follow-up (d=0.78), while Ymer et al. (2025) found CBT-SF produced significantly greater improvement in sleep quality (PSQI) and insomnia severity (ISI, d=0.95) than health education immediately post-treatment, however between group differences were not maintained at the 2 and 4 month follow-ups.

Both sleep trials received relatively high CCAT scores (Ford et al., 90%; Ymer et al., 97.5%), increasing confidence that their positive post-treatment findings were clearly reported and methodologically well described. However, the small number of sleep trials and inconsistent maintenance of effects at follow-up mean that the evidence remains preliminary.

Fatigue

Raina et al. (2022) found that the MAX self-management fatigue intervention resulted in significantly greater reductions in fatigue impact, severity and experience than health education at 8-week follow-up (d=0.61-0.87). Ymer et al. (2025) found no significant between group difference in fatigue severity at any time point; however, CBT-SF showed earlier within group improvements in fatigue than health education.

Raina et al. (2022) also received a high CCAT score (95%); however, as a small pilot RCT with a short follow-up period, its finding should be interpreted as promising rather than conclusive.

Mood, Memory-related difficulties and post-concussion symptoms

Findings were mixed for other psychological outcomes. Ymer et al. (2025) found larger within group improvements in mood, self-efficacy and quality of life for CBT-SF, but neither treatment was clearly better than the other. Rioux et al. (2024) found CBT and cognitive rehabilitation showed similar improvements in memory-related difficulties, with no evidence CBT was more advantageous (d=-0.05). Silverberg et al. (2022) found that graded exposure therapy showed significantly greater reductions in avoidance behaviour than a pacing comparator (d=0.81), while the comparator showed a small, non-significant benefit for endurance behaviour (d=0.39). Fisher et al. (2021) did not report between group efficacy data, instead, focusing on the feasibility, safety and acceptability of remote delivery.

Although Rioux et al. (2024) and Silverberg et al. (2022) received high CCAT scores (95% and 97.5%), both were feasibility focused studies with relatively small samples and active comparators. Their findings therefore provide useful preliminary evidence regarding acceptability and potential treatment effects, but do not suggest superiority of CBT-derived interventions for functional cognitive or post-concussion symptoms. Fisher et al. (2021), which received the lowest CCAT score (82.5%), primarily reported feasibility and acceptability data from a remote delivery sub-study; it therefore contributes limited evidence regarding treatment effectiveness for depression.

Feasibility and Acceptability

Across the five studies reporting on feasibility, recruitment, retention, adherence and treatment credibility were all consistently high. Retention ranged from 85% (Raina et al., 2022) to 100% (Rioux et al., 2024), and treatment was rated as credible by the 91.7%-96% of participants in the two studies that measured and reported this (Rioux et al., 2024; Silverberg

et al., 2022). Fisher et al. (2021) reported high satisfaction with remote safety monitoring protocols. Therapist fidelity to treatment manuals was consistently high (92-97% of essential content that was delivered) where reported. Two studies (Rioux et al. 2024; Silverberg et al. 2022) used the same therapists to deliver both treatment arms, which carries the risk of contamination between conditions. Rioux et al. (2024) found evidence of this in around 10% of audited sessions, where therapists delivered content more typical of the non-assigned treatment, and reported that therapists found it difficult to switch between approaches. Silverberg et al. (2022) found contamination occurred only rarely (in 6 of over 170 audited sessions), but both studies recommended that future trials use different therapists for each treatment arm to reduce this risk.

Methodological and Reporting Quality Appraisal

Lead reviewer CCAT scores ranged from 33/40 (82.5%; Fisher et al., 2021) to 39/40 (97.5%; Silverberg et al., 2022; Ymer et al., 2025), indicating generally strong methodological and reporting quality across the included studies according to the CCAT framework; scores are presented alongside study characteristics in Table 2. As the CCAT is a broad quality appraisal tool rather than a formal risk of bias instrument, these scores should be interpreted as a structured appraisal across its eight domains rather than as stand-alone estimates of risk of bias. Notably, there did not appear to be any clear relationship between CCAT score and trial outcomes.

A second reviewer independently appraised all six studies; agreement was close for five studies (within 1-2 points), and the greatest discrepancy was for Fisher et al. (2021), which received 37/40 (92.5%) from the second reviewer). This difference arose in the design and sampling domains, reflecting differing interpretations of how the CCAT criteria can be applied to a feasibility study. The lead reviewer rated these domains more conservatively (3/5 each), reflecting the paper's focus on a feasibility subsample from an ongoing pilot RCT and more limited methodological reporting relevant to these CCAT domains.

Further, common limitations across studies included relatively brief reporting of ethical procedures and, in the smaller pilot / feasibility trials, limited detail regarding randomisation and allocation concealment. Full CCAT ratings across all eight domains for both reviewers, are reported in Appendix C.

Taken together, the appraisal suggests generally good methodological and reporting quality. However, these scores were considered alongside limitations related to small pilot / feasibility samples, short follow-up periods, and variation in comparators when interpreting effects by outcome domain.

Discussion

Summary of Findings

The current review aimed to narratively synthesise evidence from six RCTs evaluating remotely delivered CBT and CBT-derived interventions for adults with ABI. The studies addressed sleep, fatigue, depression, functional cognitive symptoms, and post-concussion avoidance and endurance behaviours. Key findings indicated that remote CBT was associated with greater improvements in sleep quality and insomnia severity than the comparison groups, particularly immediately after treatment. The trial addressing fatigue also found greater reductions in fatigue in comparison to health education (Raina et al., 2022). These findings were derived from studies that received high CCAT scores, with the sleep trials scoring 90.0% and 97.5% and the fatigue trial scoring 95.0%. This supports confidence in the general methodological and reporting quality of these studies; however, it does not remove the uncertainty arising from the small number of trials, modest sample sizes, and limited follow-up periods.

Findings for depression, anxiety, memory-related concerns and quality of life were mixed, i.e., some trials found improvements across both CBT and an active comparison group as opposed to clear evidence that CBT was superior (Ymer et al., 2025; Rioux et al., 2024). Feasibility and acceptability were generally strong across the study designs and remote delivery formats, with high retention, treatment credibility and adherence when these were measured. Recruitment was also generally positive, although one study did not meet its predefined screening target (Silverberg et al., 2022).

Relation to Existing Literature

These findings are broadly consistent with wider evidence that remotely delivered and face-to-face CBT can produce comparable outcomes when directly compared (Andersson et al., 2019). However, the included ABI studies did not directly compare delivery formats. These findings also support broader evidence associated with CBT improving depression and anxiety across neurological populations, though effects for people with TBI have been smaller and less reliable than for other neurological conditions (Gandy et al., 2024).

A notable feature of the present review was the limited focus on mood outcomes within the remote CBT studies. Only one included study specifically targeted depression (Fisher et al.,

2021), and this study reported feasibility and acceptability of remote delivery rather than comparative treatment effects; no included study specifically targeted anxiety. This differs from the wider ABI and TBI literature. For example, Gandy et al. (2024) examined CBT for depression and anxiety across neurological conditions, while Seery et al. (2026) reviewed psychological interventions for depression following ABI. Both reviews included more studies targeting mood difficulties than were identified in the present review of remote CBT. However, their findings were mixed: Gandy et al. found smaller and less reliable effects for TBI than for several other neurological conditions, while Seery et al. (2026) found that the overall effect for post ABI depression became non-significant after adjustment for study limitations. This suggests that the limited mood evidence in the present review reflects a gap in research on remote delivery, while also highlighting the need for stronger evidence for mood interventions following ABI more broadly.

The effect sizes reported in the remote studies were moderate to large for some outcomes, including sleep, fatigue, and avoidance behaviour. This appears greater than the small and less reliable effects reported for mood outcomes in the wider ABI and TBI literature (Gandy et al., 2024; Seery et al., 2026). However, this comparison is limited because the studies examined different difficulties, used different outcome measures and included different comparison groups. There was some suggestion that effects from smaller pilot and feasibility studies may be less consistent in larger trials. For example, Ymer et al. (2025), the largest trial in the present review, found an immediate benefit for sleep but no clear longer-term advantage over health education, and no clear between-group benefit for fatigue. However, this pattern was not consistent across all studies, and the current evidence is too limited to conclude whether remote CBT produces different effect sizes from face-to-face CBT.

The strongest evidence was found for sleep and fatigue outcomes, though only two trials explored sleep (Ford et al., 2023; Ymer et al., 2025) and one specifically evaluated fatigue (Raina et al., 2022). Therefore, these findings should be interpreted cautiously.

The limited evidence for mood outcomes is consistent with Seery et al.'s (2026) finding, i.e., that the overall effect of psychological interventions for post ABI depression was small, and became non-significant after methodological adjustment, with non-significant effects in the TBI only subgroup. However, the current review included only a few trials exploring mood outcomes, limiting firm conclusions.

The results of the present review challenge previous concerns regarding whether cognitive difficulties may make remote interventions less accessible for people with ABI (Miao et al., 2022) as high feasibility and acceptability were generally found across the included studies. This is further supported by Bergquist et al.'s (2008) observation suggesting that people with moderate to severe TBI and memory impairment can engage successfully with internet-based interventions with appropriate support. Additionally, all interventions in the present review were facilitator-led which may have aided engagement with pacing, understanding, and use of treatment materials. This is also consistent with wider evidence showing higher completion rates for therapist guided as opposed to self-guided remote CBT (Bakanaitė et al., 2025).

Strengths and Limitations

The present review's main strength is its narrow and clearly defined focus on CBT and CBT-derived interventions delivered remotely to adults with ABI. By including RCTs only, this allowed for greater confidence in the process of considering differences between interventions and comparison groups, more so than uncontrolled studies. Additionally, methodological and reporting quality was independently appraised by two reviewers using the CCAT, providing a structured assessment across eight broad domains. Reviewer agreement was close for five of the six included studies, supporting the consistency of the appraisal process. Both reviewers' ratings and any areas of disagreement were reported transparently.

However, several limitations should also be noted. Firstly, only six RCTs met the inclusion criteria, and several were small, single site pilot or feasibility studies and were not powered to detect differences in clinical outcomes. Findings comparing the interventions should therefore be interpreted cautiously. Secondly, comparison groups varied i.e., ranging from waitlist or treatment as usual to active comparators e.g., health education, cognitive rehabilitation and pacing, limiting direct comparison of effect sizes across studies.

Thirdly, follow-up periods were also short, i.e., ranging from immediately after treatment to six months. The longer-term maintenance of treatment benefits therefore remains uncertain. This was particularly evident for sleep, as improvements were maintained at follow-up in Ford et al. (2023) but not in Ymer et al. (2025).

Finally, restricting eligibility to RCTs and English language publications may have excluded relevant non-randomised feasibility studies and research published in other languages.

Although this restriction strengthened the focus on comparative effectiveness, it may have narrowed the available evidence regarding implementation and accessibility.

Clinical and Research Implications

In terms of clinical practice, the findings suggest that facilitator-led remote CBT is a feasible and acceptable option for psychological care for adults with ABI who perhaps experience barriers to accessing face-to-face treatment or prefer remote modalities. The evidence from the present review is most encouraging for sleep, with limited evidence for fatigue. However, evidence for mood and cognitive outcomes is limited.

Good engagement was reported across videoconferencing and blended eHealth approaches, which were specifically included in the current review rather than telephone interventions. Interventions were also delivered by clinical psychologists, neuropsychologists, occupational therapists and supervised psychology students and graduates, suggesting that remote interventions may be delivered by different appropriately trained facilitators, however, the small evidence base from the present review means that conclusions regarding facilitators should remain cautious. Further, appropriate training, supervision and clear treatment manuals are important factors.

A useful next step for research would be large, multi-site trials directly comparing the same ABI adapted CBT intervention delivered by videoconference with face-to-face delivery. Such studies could examine whether remote CBT achieves clinical outcomes that are broadly comparable to face-to-face CBT, while potentially improving access for people affected by fatigue, mobility, transport and geographical barriers. Longer-term follow-up would help establish whether any benefits are maintained over time. In addition to symptom change, future trials could assess treatment completion, therapeutic alliance, treatment fidelity, patient preference, and digital accessibility. Randomisation could take into account injury type and severity, cognitive impairment, baseline symptom severity and digital access, to help identify which people are most likely to benefit from remote, face-to-face or hybrid delivery.

Future trials would also benefit from clearer reporting of ethical procedures, randomisation and allocation concealment. Where two active interventions are compared, using separate trained facilitators for each treatment arm may reduce the risk of treatment contamination identified in Rioux et al. (2024) and Silverberg et al. (2022).

Sleep and fatigue remain appropriate targets for further research, as they were the only outcomes for which the present review identified relatively consistent positive findings. However, this should not lead to mood outcomes being deprioritised. Given the prevalence and impact of depression and anxiety after ABI, future research could evaluate ABI adapted remote CBT that specifically targets these difficulties and reports comparative clinical outcomes. Research could also examine whether interventions addressing overlapping sleep, fatigue and mood difficulties are more useful for some people with ABI than interventions targeting a single symptom.

Conclusion

Remotely delivered CBT appears feasible and acceptable for adults with ABI, with encouraging evidence for improvements in sleep and some evidence for fatigue. Evidence for effectiveness on mood and cognitive outcomes remains limited. However, given the fatigue, mobility and transport barriers that can often restrict access to face-to-face psychological care following ABI, these findings support continued development and evaluation of remotely delivered CBT, alongside larger trials examining longer-term effectiveness.

References

Andersson, G., Titov, N., Dear, B. F., Rozental, A., & Carlbring, P. (2019). Internet-delivered psychological treatments: From innovation to implementation. *World Psychiatry, 18*(1), 20-28. https://doi.org/10.1002/wps.20610

Avramovic, P., Rietdijk, R., Attard, M., Kenny, B., Power, E., & Togher, L. (2023). Cognitive and behavioral digital health interventions for people with traumatic brain injury and their caregivers: A systematic review. *Journal of Neurotrauma, 40*(3–4), 159-194. https://doi.org/10.1089/neu.2021.0473

Bakanaitė, R., Bakanienė, I., & Kanyeredzi, A. (2025). Long-term effectiveness of internet-based versus traditional cognitive behavioral therapy across psychiatric disorders: A systematic review and meta-analysis. *Telemedicine and e-Health, 31*(11), 1284-1296. https://doi.org/10.1177/15305627251358855

Baylan, S., Griffiths, S., Grant, N., Broomfield, N. M., Evans, J. J., & Gardani, M. (2020). Incidence and prevalence of post-stroke insomnia: A systematic review and meta-analysis. *Sleep Medicine Reviews, 49*, 101222. https://doi.org/10.1016/j.smrv.2019.101222

Bergquist, T., Gehl, C., Lepore, S., Holzworth, N., & Beaulieu, W. (2008). Internet-based cognitive rehabilitation in individuals with acquired brain injury: A pilot feasibility study. *Brain Injury, 22*(11), 891-897. https://doi.org/10.1080/02699050802405487

Bulloch, A. G. M., Fiest, K. M., Williams, J. V. A., Lavorato, D. H., Berzins, S. A., Jetté, N., Pringsheim, T. M., & Patten, S. B. (2015). Depression-A common disorder across a broad spectrum of neurological conditions: A cross-sectional nationally representative survey. *General Hospital Psychiatry, 37*(6), 507-512. https://doi.org/10.1016/j.genhosppsych.2015.06.007

Burgess, C., Tian, E. J., Tyack, E., & Kumar, S. (2024). Barriers and enablers to physical activity for individuals living with traumatic brain injury: A mixed methods systematic review. *Brain Injury, 38*(13), 1157-1170. https://doi.org/10.1080/02699052.2024.2381053

Crowe, M., & Sheppard, L. (2011). A general critical appraisal tool: An evaluation of construct validity. *International Journal of Nursing Studies, 48*(12), 1505-1516. https://doi.org/10.1016/j.ijnurstu.2011.03.004

Crowe, M., Sheppard, L., & Campbell, A. (2012). Reliability analysis for a proposed critical appraisal tool demonstrated value for diverse research designs. *Journal of Clinical Epidemiology, 65*(4), 375-383. https://doi.org/10.1016/j.jclinepi.2011.08.006

Crowe, M. (2013). *CCAT user guide* (Version 1.4). https://conchra.com.au/wp-content/uploads/2015/12/CCAT-user-guide-v1.4.pdf

Crowe, M., Sheppard, L., & Campbell, A. (2011). Comparison of the effects of using the Crowe Critical Appraisal Tool versus informal appraisal in assessing health research: A randomised trial. *International Journal of Evidence-Based Healthcare, 9*(4), 444-449. https://doi.org/10.1111/j.1744-1609.2011.00237.x

Fisher, L. B., Tuchman, S., Curreri, A. J., Markgraf, M., Nyer, M. B., Cassano, P., Iverson, G. L., Fava, M., Zafonte, R. D., & Pedrelli, P. (2021). Transitioning from in-person to remote clinical research on depression and traumatic brain injury during the COVID-19 pandemic: Study modifications and preliminary feasibility from a randomized controlled pilot study. *JMIR Formative Research, 5*(12), e28734. https://doi.org/10.2196/28734

Ford, M. E., Geurtsen, G. J., Groet, E., Rambaran Mishre, R. D., van Bennekom, C. A. M., & van Someren, E. J. W. (2023). A blended eHealth intervention for insomnia following acquired brain injury: A randomised controlled trial. *Journal of Sleep Research, 32*(1), e13629. https://doi.org/10.1111/jsr.13629

Gandy, M., Woldhuis, T., Wu, W., Youssef, M., Bisby, M. A., Dear, B. F., Heriseanu, A. I., & Scott, A. J. (2024). Cognitive behavioral interventions for depression and anxiety in adults with neurological disorders: A systematic review and meta-analysis. *Psychological Medicine, 54*(12), 3237-3250. https://doi.org/10.1017/S0033291724001995

GBD 2019 Stroke Collaborators. (2021). Global, regional, and national burden of stroke and its risk factors, 1990–2019: A systematic analysis for the Global Burden of Disease

Study 2019. *The Lancet Neurology, 20*(10), 795-820. https://doi.org/10.1016/S1474-4422(21)00252-0

Hart, T., Vaccaro, M., Krasucki, L., Chervoneva, I., & Rabinowitz, A. (2026). Increasing physical activity in moderate-severe traumatic brain injury: Protocol for a two-stage randomized controlled trial of a remote, mHealth-enhanced intervention. *Frontiers in Rehabilitation Sciences, 7*, 1656326. https://doi.org/10.3389/fresc.2026.1656326

Mar, J., Arrospide, A., Begiristain, J. M., Larrañaga, I., Elosegui, E., & Oliva-Moreno, J. (2011). The impact of acquired brain damage in terms of epidemiology, economics and loss in quality of life. *BMC Neurology, 11*, Article 46. https://doi.org/10.1186/1471-2377-11-46

Mathias, J. L., & Alvaro, P. K. (2012). Prevalence of sleep disturbances, disorders, and problems following traumatic brain injury: A meta-analysis. *Sleep Medicine, 13*(7), 898-905. https://doi.org/10.1016/j.sleep.2012.04.006

Mehta, S., Peynenburg, V. A., & Hadjistavropoulos, H. D. (2019). Internet-delivered cognitive behaviour therapy for chronic health conditions: A systematic review and meta-analysis. *Journal of Behavioral Medicine, 42*(2), 169-187. https://doi.org/10.1007/s10865-018-9984-x

Miao, M., Rietdijk, R., Brunner, M., Debono, D., Togher, L., & Power, E. (2022). Implementation of web-based psychosocial interventions for adults with acquired brain injury and their caregivers: Systematic review. *Journal of Medical Internet Research, 24*(7), e38100. https://doi.org/10.2196/38100

Minozzi, S., Cinquini, M., Gianola, S., Gonzalez-Lorenzo, M., & Banzi, R. (2020). The revised Cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application. *Journal of Clinical Epidemiology, 126*, 37-44. https://doi.org/10.1016/j.jclinepi.2020.06.015

Minozzi, S., Dwan, K., Borrelli, F., & Filippini, G. (2022). Reliability of the revised Cochrane risk-of-bias tool for randomised trials (RoB 2) improved with the use of implementation instruction. *Journal of Clinical Epidemiology, 141*, 99-105. https://doi.org/10.1016/j.jclinepi.2021.09.021

Ouellet, M.-C., Beaulieu-Bonneau, S., & Morin, C. M. (2006). *Insomnia in patients with traumatic brain injury: Frequency, characteristics, and risk factors*. *Journal of Head Trauma Rehabilitation, 21*(3), 199-212. https://doi.org/10.1097/00001199-200605000-00001

Ouellet, M.-C., Beaulieu-Bonneau, S., & Morin, C. M. (2015). Sleep-wake disturbances after traumatic brain injury. *The Lancet Neurology, 14*(7), 746-757. https://doi.org/10.1016/S1474-4422(15)00068-X

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews, 5*, Article 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Raina, K. D., Morse, J. Q., Chisholm, D., Whyte, E. M., & Terhorst, L. (2022). An internet-based self-management intervention to reduce fatigue among people with traumatic brain injury: A pilot randomized controlled trial. *American Journal of Occupational Therapy, 76*(4), 7604205100. https://doi.org/10.5014/ajot.2022.048587

Rioux, M., Mamman, R., Byworth, M. T., Panenka, W. J., Howard, A. K., Perez, D. L., Schmidt, J., Courchesne, C., LeMoult, J., Heran, M. K. S., & Silverberg, N. D. (2024). Pilot feasibility randomised controlled trial of cognitive-behavioural therapy for functional cognitive disorder after concussion. *BMJ Neurology Open, 6*, e000666. https://doi.org/10.1136/bmjno-2024-000666

Scholten, A. C., Haagsma, J. A., Cnossen, M. C., Olff, M., van Beeck, E. F., & Polinder, S. (2016). Prevalence of and risk factors for anxiety and depressive disorders after traumatic brain injury: A systematic review. *Journal of Neurotrauma, 33*(22), 1969-1994. https://doi.org/10.1089/neu.2015.4252

Seery, C., Twomey, D. M., Dully, J., Breheny, E., Skelly, T., Davenport, L., Bramham, J., Brady, N., & O'Keeffe, F. (2026). Effect of psychological interventions on depressive symptoms in the long term after acquired brain injury: A systematic review and meta-analysis. *Neuropsychological Rehabilitation, 36*(2), 392-441. https://doi.org/10.1080/09602011.2025.2499745

Silverberg, N. D., Cairncross, M., Brasher, P. M. A., Vranceanu, A.-M., Snell, D. L., Yeates, K. O., Panenka, W. J., Iverson, G. L., Debert, C. T., Bayley, M. T., Hunt, C., Baker, A., & Burke, M. J. (2022). Feasibility of concussion rehabilitation approaches tailored to psychological coping styles: A randomized controlled trial. *Archives of Physical Medicine and Rehabilitation, 103*(8), 1565-1573. https://doi.org/10.1016/j.apmr.2021.12.005

Stalder-Lüthy, F., Messerli-Bürgy, N., Hofer, H., Frischknecht, E., Znoj, H., & Barth, J. (2013). Effect of psychological interventions on depressive symptoms in long-term rehabilitation after an acquired brain injury: A systematic review and meta-analysis. *Archives of Physical Medicine and Rehabilitation, 94*(7), 1386-1397. https://doi.org/10.1016/j.apmr.2013.02.013

Stroke Association. (n.d.). *Depression and anxiety after stroke*. Retrieved July 30, 2026, from https://www.stroke.org.uk/stroke/effects/emotional-and-behavioural/depression-and-anxiety

Whitnall, L., McMillan, T. M., Murray, G. D., & Teasdale, G. M. (2006). Disability in young people and adults after head injury: 5–7-year follow-up of a prospective cohort study. *Journal of Neurology, Neurosurgery & Psychiatry, 77*(5), 640-645. https://doi.org/10.1136/jnnp.2005.078246

Ymer, L., McKay, A., Wong, D., Frencham, K., Grima, N., Roper, M., Nguyen, S., Murray, J., Spitz, G., & Ponsford, J. (2025). Randomized controlled trial of cognitive behavioural therapy versus health education for sleep disturbance and fatigue following stroke and traumatic brain injury. *Journal of Rehabilitation Medicine, 57*, jrm41302. https://doi.org/10.2340/jrm.v57.41302

Zhu, H., Wen, Q., Zhang, F., Li, S., Wang, X., & Li, J. (2026). The efficacy of cognitive behavioral therapy for sleep disorders following acquired brain injury: A systematic

review and meta-analysis. *Sleep Medicine, 138*, 108678. https://doi.org/10.1016/j.sleep.2025.108678

Chapter 2 Major Research Project

Comparing face-to-face and direct-to-home approaches to neuropsychological assessment

Prepared in accordance with the author requirements for Archives of Clinical Neuropsychology, https://academic.oup.com/acn/pages/author-guidelines

Keywords: neuropsychology; cognitive assessment; teleneuropsychology; remote testing; face-to-face assessment

Plain Language Summary

This study explored whether cognitive tests give the same results when carried out remotely (by video call) compared to face-to-face, and whether this differs between younger and older adults. Remote testing has become more common in recent years, partly due to the COVID-19 pandemic, and offers a way to improve access to healthcare services, particularly for people living in remote areas such as the Scottish Highlands and Islands.

16 healthy adults took part in the current study, eight aged 18-40 and eight aged 65 and over. Each participant completed the same set of cognitive tests twice, once in person and once remotely, using different but equivalent versions of one test to avoid participants simply memorising answers. Session order was balanced so that half completed the tests in person first, and half remotely first. This meant that practice effects would be distributed evenly across groups. Cognitive tests covered areas such as memory, language, attention, executive function and inhibition.

The findings showed that for most of the tests, scores were consistent whether participants were tested in person or remotely, and this was true for both the younger and older age groups, with no overall difference in performance between the two formats for either age group. One test which measures language, specifically picture naming and word finding did not show consistent scores between the two testing conditions. The order in which participants completed the two testing formats did make some difference to scores for a few tests, with younger and older adults tending to perform better under opposite orderings.

Overall, the findings provide preliminary support for the use of remote testing for some cognitive measures in this healthy sample. However, the small sample means that larger studies, including clinical populations, are needed before firm conclusions can be drawn about whether remote and face-to-face testing produce equivalent results. It also highlights that a small number of tests, and the order in which tests are given, may need extra consideration when comparing remote and in-person results. This research adds to a small but growing evidence base on remote cognitive testing and may help inform how services deliver assessments to people who cannot easily attend in person.

Abstract

As remote video-based cognitive assessment becomes more widely used in clinical practice, there is a gap in the evidence base raising questions whether it produces results equivalent to traditional face-to-face assessment, particularly for cognitive domains beyond verbal tasks. 16 healthy adults (8 age 18-40 and 8 aged 65 and over) completed a battery of neuropsychological tests covering a variety of cognitive domains including: verbal fluency, language, attention, processing speed, executive function, working memory, inhibition, and visuospatial abilities. Participants completed the cognitive assessment once face-to-face and once remotely, in a counterbalanced order. Consistency between the two formats, measured using intraclass correlation coefficients varied across measures: four measures (the Oral Trail Making Test A, RBANS Total Scale Score, DKEFS Colour Naming, and RBANS Attention Index) show excellent consistency; several memory and executive function measures showed only fair consistency, and the RBANS Language Index showed poor consistency. Follow-up analysis for the RBANS Language Index explored underlying subtests and found no conclusive explanation. Overall, test scores did not differ significantly between face-to-face and remote formats, with no significant age-related differences found either. A more unexpected finding was that the order in which participants completed the two formats interacted with age for several executive function measures: younger participants (18-40) tended to perform slightly better having done the face-to-face session first, while older participants (65 and over) performed better having done the remote session first. Overall, these preliminary findings suggest that scores were broadly comparable across remote and face-to-face conditions for several measures in this healthy sample; however, the small sample size and limited statistical power mean that equivalence between formats cannot be established. The RBANS Language Index showed poor consistency between conditions in this sample; however, this isolated finding may reflect chance variation and requires replication in a larger sample. Larger studies could also examine the age-related order effect.

Introduction

History of remote cognitive testing

Remote cognitive assessment has been used for a considerable time, particularly to help improve access to neuropsychology for people living in remote, sparsely populated areas (Singh & Germine, 2021). The COVID-19 pandemic significantly accelerated the adoption of remote cognitive testing, and remote ways of working became relevant to many more people and as such research on this topic gained interest and pace (Fox-Fuller et al., 2021; Bilder et al., 2020). Remote cognitive testing became a critical component of telehealth services, offering patients the option to undergo cognitive assessments from home, whilst maintaining regular contact with healthcare providers. Since the pandemic, there has been continued interest in remote assessment as well as hybrid approaches, incorporating both traditional face-to-face and technology-based modalities.

Decades of literature supports the feasibility of traditional paper-pencil cognitive tests and in comparison to this traditional modality, remote administration is much more recent (Schmand, 2019). Although neuropsychological testing has remained heavily reliant on paper-pencil-tests (Van Patten, 2021), new platforms such as Q Global, (Pearson, 2016) have enabled progress in remote testing. Yet, there remains a notable lack of direct comparisons between remote and face-to-face cognitive formats (Sperling et al., 2023), making it difficult to fully evaluate the equivalence of these methods.

Further, certain cognitive domains, such as processing speed and executive function, pose more challenges when assessed in remote formats. Remote measurement in these domains can be difficult due to the need for manipulation of materials, manual response formats and precise measurement of response times, alongside environmental variability, concerns over test integrity and test condition control (e.g., ensuring a standard test environment free from distractions and a good internet connection). These challenges can potentially impact the feasibility and acceptability of remote cognitive testing as well as its accuracy and reliability. Further research is needed to establish more robust comparisons and to identify ways to overcome challenges specific to remote testing of these cognitive domains, e.g., using technology to minimise potential timing errors, and ways of ensuring standardised procedures are followed.

In addition, remote cognitive testing can be influenced by age in several ways. For example, comfort with technology, accessibility and familiarity with digital devices vary across age

groups, potentially making remote testing more accessible to younger individuals (Anderson & Perrin, 2017). Older adults may experience greater difficulties with remote platforms, which could amplify age-related effects in cognitive domains, such as processing speed (Anderson & Perrin, 2017).

Research on age effects in remote cognitive testing is mixed. While some studies report minimal differences between remote and face-to-face testing for older adults (Cyr et al., 2021), others highlight adaptation challenges for this demographic (Belleville et al., 2023). Age may therefore interact with remote cognitive testing in terms of both cognitive performance and test accessibility.

The rapid development of remote cognitive testing in recent clinical practice has meant scientific research has not kept pace. This gap has led to many remote tests being available, yet lacking robust empirical support, raising concerns about their reliability (Parsons, 2016). This highlights the need for more research on the remote delivery and evaluation of cognitive tests, focusing on identifying the most reliable and robust tests, particularly to address accessibility challenges faced in rural areas, such as NHS Highland.

Acceptability of remote assessment

Research to date supports the acceptability of remote cognitive assessment (Parsons et al., 2022; Sumpter et al., 2022; Gardner et al., 2021; Wadsworth et al., 2018; Cullum et al., 2014; Parikh et al., 2013) and the benefits for patients. For example, remote delivery addresses logistical and financial barriers (e.g., transportation costs, work and childcare commitments) which can be associated with travelling to attend a face-to-face appointment (Van Patten, 2021). Further, the travel involved with attending face-to-face appointments presents concern for the safety of vulnerable individuals with cognitive impairments due to symptoms of confusion, balance impairment, and difficulty navigating directions (Sperling et al., 2023). These elements enhance the appeal of remote cognitive testing. Clinicians often find remote cognitive testing acceptable as it allows for more efficient use of their time and resources, e.g., eliminating travel time associated with home visits. Consequently, remote options can help address barriers to accessing cognitive assessments by providing flexible input options for service users (Sumpter et al., 2022). Acceptability of remote cognitive testing has been demonstrated in an increasing number of studies. For example, Parsons and colleagues (2022) reported that 72 patients (paediatric to older adult age range) were indeed satisfied with the remote modality, with clinicians reporting that session goals were achieved in 90%

of remote contacts. Singh and colleagues (2021) looked at remote cognitive testing using a clinical sample and found test acceptability for measures of processing speed and working memory. Despite these encouraging results, gaps in the literature remain, e.g., in considering how acceptability may vary across specific clinical populations, the specific testing methods used and the context in which the testing is conducted. Further, the acceptability of remote cognitive testing can be influenced by the participant's comfort level with technology and therefore providing adequate support for those who may struggle with technology is fundamental to maintaining high levels of acceptability.

Feasibility, reliability, and validity of remote assessment

When evaluating cognitive testing, particularly in remote settings, it is important to consider its feasibility, reliability, and validity. These factors help ensure that cognitive assessments are effective, accurate and dependable. With reference to feasibility, i.e., the practicalities of implementing remote cognitive testing, there are limitations, which can subsequently impact the validity (i.e., the extent to which a test measures what it is intended to measure) and reliability (i.e., the consistency of test results over time and across contexts). For example, additional equipment may need to be purchased, a steady internet connection is crucial (Jaffe et al., 2020), and a private space at the individual's home is essential to minimise disturbances. All these factors may mean that the environment in which tests are administered differs from that which they were designed to be administered within (Barney et al., 2020). Moreover, background noise and interruptions to audio/visual continuity can pose threats to test validity and reduce standardisation of the testing environment (Sumpter et al., 2022). Therefore, test instructions need to be clearly outlined, and considerations taken into account to limit these potential threats. This pattern is also reflected in wider guidance and evidence on remote testing. UK guidance from the British Psychological Society (2020) states that assessments based on simple verbal question and answer formats are among the most suitable for remote administration, compared to those that involve handling and administering test materials remotely. This is supported by a systematic review and meta-analysis of videoconference administered neuropsychological tests (Brearly et al., 2017), which found that verbally based tasks including digit span, verbal fluency, and word list learning, were not affected by remote administration, whereas tests involving physical handling of materials produced more mixed results across studies. More specifically, one study (Galusha-Glasscock et al., 2016) investigated the feasibility and reliability of the RBANS (Randolph, 1998) by comparing video teleconference (remote) with face-to-face test

conditions, finding significant correlations between conditions across all index scores (ranging from r = .59 to r = .90) and no statistically significant differences between the mean scores obtained in each condition. However, the small sample size of 18 participants makes it difficult to generalise findings. It is also noted that participants in this study were recruited from an Alzheimer's Disease Centre connected to a university and were a clinical sample consisting of a mix of cognitively normal controls, individuals with a diagnosis of Alzheimer's and individuals with mild cognitive impairment (MCI) known to the research centre. Additionally, the remote condition was completed on site with site Wi-Fi. Therefore, with only one preliminary experimental study supporting the remote administration of the RBANS, more research is needed to test the remote condition from home and with a different population. More broadly, the evidence base for remote versus in-person cognitive assessment remains concentrated on a small number of cognitive domains, predominantly verbal tasks, leaving other cognitive domains such as processing speed, executive function and inhibition, visuospatial/constructional abilities, attention, and working memory comparatively under examined. The present study aims to contribute to this wider evidence base by assessing a broader range of cognitive domains within the same design.

Current Study Aims

The current study directly compared remote and face-to-face cognitive testing formats in healthy younger and older adult groups, using the RBANS alongside measures of attention, processing speed, executive function, inhibition, verbal fluency, and working memory. The healthy sample was chosen rather than a clinical one as a preliminary step in establishing this research in the Highland context. The measures were selected due to their being most frequently used in our clinical practice, applicability across a wide range of neuropsychological assessments, and feasible to administer within approximately 45 minutes, to minimise participant burden.

The study focused on the reliability and comparability of neuropsychological test scores obtained from face-to-face and remote assessments across younger and older adults, and aimed to address the following:

1. Primary hypothesis: Remote and face-to-face scores will show moderate to high consistency across cognitive measures.

2. Secondary hypothesis: Consistency between remote and face-to-face scores will be lower for older adults than younger adults.

3. Exploratory question: Do test scores differ according to administration condition or age group across the cognitive measures assessed?

Method

Study Design

A mixed factorial experimental design was used to examine performance of younger and older adults [between subjects factor] with two levels: younger (18-40 years), older (60+ years) on tests covering a range of cognitive domains including attention, language, visuospatial/constructional abilities, immediate and delayed memory, processing speed, and executive function across two administration modes [within subjects]: remote, face-to-face, with counterbalancing of administration order between subjects (face-to-face first, remote first). Although not all measures used had parallel forms available, these were used where available, again with order counterbalanced between subjects (version A first, version B first). Crossing of the experimental factors led to four potential testing sequences, i.e. in-person version A then remote version B; in-person version B then remote version A; remote version A then in-person version B; remote version B then in-person version A. Participants were randomly allocated to one of these four predefined sequences using randomizer.org (Urbaniak & Plous, 2019), with allocation balanced across age groups. Participants were first assigned to age group (younger, older) based on predefined criteria (aged 18-40, aged 60+ years).

Ethical Approval

Approval was sought and obtained from the University of Glasgow College of Medical, Veterinary and Life Sciences (MVLS) Ethics Committee (200250084); (see Appendix F).

Participants

A community sample was recruited via posters, email lists and social media associated with local community organisations (e.g., community groups and outdoor activity clubs). Participants self-selected by responding to advertisements Recruitment also occurred via snowball sampling, through word-of-mouth sign-up.

Upon expressing interest, potential participants were sent an information sheet and consent form (Appendix G, participant information sheet; Appendix H, Informed consent form). Written informed consent was secured prior to participation, and the right to decline or withdraw without consequence was emphasised. Participants attended both a face-to-face appointment and a remote appointment (each ~1-hour duration).

The original target sample size was 20 participants (10 per age group), set on pragmatic grounds reflecting what could practically be achieved within the time available to complete

the study. However, the counterbalanced design required multiples of 8 participants (4 administration sequences for each age group), hence the target sample size was amended to 16 participants. This decision was taken after ethical approval had been obtained, but prior to the start of recruitment. It was not expected to meaningfully impact on statistical power and would still produce a useful dataset to enable preliminary analysis and inform the conduct of future studies.

Inclusion and Exclusion Criteria

Participants were healthy adults within one of two age brackets: 18-40 years or 65 years and over. Participants were required to be based in Inverness or Glasgow, or willing to travel to the face-to-face appointment. Participants also needed access to a laptop, a stable internet connection, and a private space to complete the remote condition.

Demographic Characteristics of Participants

16 participants took part in the study, eight younger and eight older. The overall sample had a mean age of 50.94 years (SD = 20.23, range = 21–73) and a median age of 52.5 years. Eleven participants were female (68.8%) and five male (31.3%). Years of formal education ranged from 11 to 27 years.

Given age group was a between-subjects factor of interest, gender and education were checked for equivalence between the younger and older groups before proceeding to the main analyses. Fisher's Exact test indicated that gender distribution did not differ significantly between the younger and older groups (p = .282). However, a Mann-Whitney U test showed that education differed between the younger and older groups with younger participants reporting more time in education (Mdn = 19.00) than older participants (Mdn = 13.00), U = 12.00, z = -2.13, p = .038. The variable years of education was hence included as a covariate in the secondary analysis.

Table 3 *Demographic characteristics of participants*

Variable	Younger (n = 8)	Older (n = 8)	Total (N = 16)
Age, M (SD)	31.88 (6.38)	70.00 (2.45)	50.94 (20.23)
Gender, n (% female)	7 (87.5%)	4 (50.0%)	11 (68.8%)
Education (years), Mdn (range)	19.00 (16–20)	13.00 (11–27)	17.00 (11–27)

Note. Education differed significantly between age groups, $U = 12.00$, $z = -2.13$, $p = .038$

Cognitive Measures

Cognitive tests were administered in line with copyright restrictions and test security permissions (Pearson, 2020) as well as BPS guidelines (British Psychological Society, 2021). Parallel versions were used where available to minimise practice effects, particularly for RBANS memory and naming items where repeated exposure to identical stimuli could inflate retest performance. Hence two parallel forms of the RBANS (Form A and Form B) were used across the two administration sessions, with form order counterbalanced alongside administration mode, and Digit Span (Forward and Backward) used different trial sets across the two sessions (using sequences given in Lezak et al., 2012). Parallel forms were not available for verbal fluency (FAS and Animal Naming), the Oral Trail Making Test, or the DKEFS Colour-Word Interference Test, so identical items were used on both testing sessions for these measures. Practice effects based on exposure to specific test items were considered somewhat less relevant for these tasks given their format.

Remote administration followed standardised face-to-face procedures as closely as possible. For remote sessions, materials were shared via screen share using Q Global, and participants were asked to ensure they were in a private space for the duration of the appointment. Participants were informed that sessions were not to be recorded. For drawn test responses, the researcher destroyed materials following face-to-face sessions; for remote sessions, participants were observed to destroy their own drawn material on screen during the session, with the exception of materials required for scoring, which were posted back to the researcher (following Sumpter et al., 2022). Remote sessions were delivered via Microsoft Teams.

Table 4 *Cognitive Measures*

Test	Cognitive domain	Description / Rationale	Time to administer	Adaptations for remote delivery
FAS + animal naming (Tombaugh et al., 1999)	Verbal fluency Processing Speed Executive function	Well established measure of language and EF commonly used in remote assessments	Approx. 5 minutes in total	None. Easy to administer remotely using verbal instructions
RBANS forms A&B (Randolph, 1998)	Immediate memory Delayed memory Visuospatial / constructional abilities Language Attention	RBANS is an efficient, reliable tool that assesses several cognitive domains, making it ideal for screening in both face-to-face and remote settings. Parallel forms (A and B) were used given the risk of practice effects with recall-based stimuli	Approx. 20-30 minutes	Coding to be sent in a sealed envelope
Oral trails (Mrazik et al., 2010)	Processing Speed Executive function Attention	Another well-established measure well suited to remote administration. Sequence of verbal tasks that involve connecting numbers/letters	Approx. 5 minutes	None. Does not require visual or motor demands required in written tests

Digit Span Forward and Backward (version from Lezak, 1995 3rd Edition)	Working memory Executive function Attention	Assesses cognitive control, short-term memory. Different trial sets were used across sessions to minimise practice effects	Approx. 5 minutes	None. Simple to administer remotely; requires verbal response and recording only
DKEFS Colour Word Interference (Delis et al., 2001)	Processing Speed Executive Function Inhibition	A comprehensive test, providing insight into executive function, a cognitive domain crucial for daily decision-making and problem solving	Approx. 10-15 minutes	Can be administered remotely via Q Global with screen shared instructions

Research Procedure

Participants completed the study with the researcher present either at a face-to-face location or remotely. Prior to testing, participants were provided with instructions and had the opportunity to ask questions. Written or electronic informed consent was obtained prior to or at the first appointment, after which cognitive testing was administered for that session. Participants then completed cognitive testing in the alternate condition at a second appointment, scheduled at least two weeks after the first. A two-to-four-week interval was considered the most feasible option given pragmatic constraints. Similar intervals have been used in the neuropsychology literature, e.g., Bartels and colleagues (2010) used a 2–3-week interval between initial testing sessions, while Bergman and colleagues (2025) used a four-week interval, both to balance the minimisation of practice effects against pragmatic constraints. Practice effects are well documented across this kind of timeframe (Collie et al., 2003), highlighting the importance of the counterbalanced design adopted in the present study to minimise this risk.

The total duration of the study was approximately 9 months, including ethical approval, recruitment, and testing.

Analysis Plan

Data preparation involved checking for missing data, outliers, and inconsistencies, and assessing the distribution of scores more generally using descriptive statistics including measures of skewness and kurtosis. Intraclass Correlation Coefficients (ICC) were calculated across the whole sample to assess the reliability of test scores, i.e., to ascertain whether scores obtained via the two administration methods were comparable by measuring the consistency of individual scores across face-to-face and remote administration. ICC values were interpreted in line with Cicchetti's (1994) guidelines, where values below .40 indicate poor reliability, .40-.59 fair reliability, .60-.74 good reliability, and .75-1.00 excellent reliability. Spearman's rank-order correlation coefficient was also calculated between face-to-face and remote scores for the whole sample, as a non-parametric check of the ICC findings.

An exploratory repeated-measures analysis of variance was used to examine whether test performance differed significantly according to administration mode (face-to-face vs remote), age group (younger vs older), or testing order, across the 16 primary cognitive measures, with years of education included as a covariate. Multivariate test statistics (MANOVA) were used to analyse the within-subjects effect of administration mode. Due to the small sample size

relative to the number of dependent variables, multivariate test statistics could not be calculated for the between-subjects effects (residual degrees of freedom = 11, against 16 dependent variables); univariate tests are reported instead. This allowed for a preliminary assessment of whether performance differed across administration conditions, age groups, or testing order, beyond the correlational analyses. Given limited power (see Sample Size and Power, below), the emphasis was placed on effect size rather than statistical significance alone: partial eta squared (η^2p) was reported alongside all F statistics, interpreted using Cohen's (1992) thresholds (small = .01, medium = .06, large = .14).

Sample Size and Power

G*Power (Faul et al., 2007) was used to conduct sensitivity power analyses to determine the smallest effects the study would be able to detect with adequate power, for both the planned and achieved sample sizes.

For the correlational analyses, with a planned sample size of 20 participants, the study had 80% power to detect a correlation of $r = .579$ or larger at a significance level of $\alpha = .05$ (see Appendix L). A repeated sensitivity analysis based on the achieved sample of 16 participants indicated that the study had 80% power to detect a correlation of $r = .636$ or larger at the same significance level (see Appendix L), meaning the study was less able to detect smaller effects than originally planned. However, the study was sufficiently powered to detect strong correlations of the sort that would be indicative of 'good' reliability (.6-.74) or excellent reliability (.75-1), according to Cicchetti's (1994) guidance.

A similar sensitivity analysis was conducted for the MANOVA, based on the five cognitive tests used (Verbal fluency, RBANS, Oral Trail Making Test, Digit Span and the DKEFS Colour Word Interference), which together comprised the 16 primary dependent variables analysed in the final dataset. With a planned sample size of 20, the minimum detectable effect size, f^2, was 0.538, indicating the study was powered to detect only large effects (see Appendix L). Based on the achieved sample of 16, the minimum detectable effect size increased to $f^2 = 0.738$ (see Appendix L), indicating that only very large effects could be reliably detected at adequate power. Of course, this secondary analysis was exploratory in nature, and other studies in this area have recruited comparable samples (Galusha-Glasscock et al., 2016).

Results

Preliminary Analyses

Variable Selection

Prior to analysis, the research team identified a set of primary variables from the numerous possible metrics available, to reduce the number of calculations and risk of false positives. Primary variables formed the focus of the analyses reported below, and comprised Verbal Fluency (FAS total, Animal total), RBANS index scores (Immediate Memory, Visuospatial/Constructional, Language, Attention, Delayed Memory, Total Scale Score), Trail Making Test A and B (z-scores), Digit Span (forward and backward raw scores), and DKEFS Colour Word Interference (Colour Naming, Word Reading, Inhibition, Inhibition Switching). Supplementary measures were not analysed in the present study, as the primary variables provided sufficient coverage of the cognitive domains relevant to the research aims. The primary variables were selected before the study data were examined and were not selected on the basis of the observed results. The power calculation related to the planned correlation and multivariate analyses, rather than to each individual cognitive variable separately.

Assessment of Normality

All primary cognitive variables were assessed for normality within each age group using boxplots, skewness, and kurtosis values, and the Shapiro-Wilk test (appropriate given the small per-group sample size, n = 8).[1] Given the small sample size and the presence of non-normal distributions across several variables, a non-parametric approach was adopted throughout the primary analyses, consistent with recommendations for small or non-normally distributed samples (Field, 2018). Descriptive statistics are therefore reported using medians rather than means, and Spearman's rank-order correlations were used alongside the Intraclass Correlation Coefficient as a non-parametric check of consistency.

[1] The majority of primary variables did not significantly deviate from normality in either age group ($p > .05$). Of the 64 Shapiro-Wilk tests conducted (16 primary measures x 2 conditions x 2 age groups), 8 (12.5%) were statistically significant, including: Digit Span Backward at In-person administration (Younger, p = .004); Digit Span Forward at Remote administration (Younger, p = .007) and In-person administration (Older, p = .025); the Word Reading subtest at In-person administration (Younger, p = .029) and Remote administration (Older, p = .048); TMT-B (z) at In-person administration (Older, p = .003); the Delayed Memory Index at Remote administration (Younger, p = .026); and the Visuospatial Index at Remote administration (Older, p = .009).

Descriptive Statistics for Primary Cognitive Variables

Median, interquartile range (IQR), minimum, and maximum scores for each primary cognitive variable are reported in Table 5, separately for in-person and remote administration, and split by age group.

Table 5 *Median (IQR range) for primary cognitive variables by administration condition and age group*

Measure	Younger Mdn (IQR)	Older Mdn (IQR)	Total Mdn (IQR)
FAS In-Person	37.50 (17.25)	45.50 (12.75)	42.00 (16.50)
FAS Remote	41.00 (23.00)	46.50 (24.50)	46.50 (22.00)
Animals In-Person	24.00 (3.75)	21.00 (6.25)	22.50 (4.75)
Animals Remote	22.50 (6.50)	22.00 (10.00)	22.50 (8.00)
Immediate Memory In-Person	100.00 (14.25)	103.00 (26.25)	100.00 (19.50)
Immediate Memory Remote	104.50 (29.00)	104.50 (7.50)	104.50 (12.80)
Visuospatial In-Person	94.50 (27.75)	109.00 (33.50)	103.50 (31.00)
Visuospatial Remote	107.00 (43.00)	103.50 (12.00)	103.50 (29.50)
Language In-Person	108.00 (24.50)	97.50 (10.75)	100.00 (19.75)
Language Remote	96.00 (22.25)	97.00 (17.50)	97.00 (18.25)
Attention In-Person	94.00 (19.50)	103.00 (22.50)	100.00 (17.25)
Attention Remote	98.50 (19.50)	109.00 (20.25)	100.00 (17.25)
Delayed Memory In-Person	100.00 (17.00)	100.50 (17.50)	100.50 (16.75)
Delayed Memory Remote	100.00 (40.00)	101.00 (17.25)	101.00 (18.25)
Total Scale In-Person	99.00 (13.75)	103.50 (23.25)	99.00 (22.00)

Total Scale - Remote	99.00 (21.50)	104.50 (12.75)	102.50 (10.50)
TMT-A (z) In-Person	-1.72 (3.76)	-0.61 (0.84)	-0.87 (1.75)
TMT-A (z) Remote	-1.71 (4.50)	-0.72 (0.44)	-0.81 (1.36)
TMT-B (z) In-Person	-0.95 (1.61)	0.62 (0.98)	0.01 (1.76)
TMT-B (z) Remote	-0.42 (0.84)	0.53 (0.79)	0.05 (1.09)
Digit Span Forward In-Person	9.00 (2.75)	9.00 (2.50)	9.00 (2.75)
Digit Span Forward Remote	8.50 (1.00)	9.50 (2.76)	9.00 (2.00)
Digit Span Backward In-Person	8.00 (1.75)	7.50 (3.25)	8.00 (1.75)
Digit Span Backward Remote	7.00 (1.75)	7.50 (2.00)	7.00 (2.00)
Colour Naming In-Person	9.50 (3.00)	10.00 (5.25)	10.00 (3.00)
Colour Naming Remote	8.00 (3.50)	9.50 (6.50)	8.50 (4.75)
Word Reading In-Person	11.00 (4.00)	10.50 (2.50)	11.00 (2.50)
Word Reading Remote	10.50 (2.00)	11.00 (3.50)	11.00 (2.75)
Inhibition In-Person	11.00 (4.25)	11.50 (2.75)	11.00 (2.75)
Inhibition Remote	11.50 2.50)	11.00 (4.00)	11.00 (2.50)
Inhibition Switching In-Person	9.00 (4.25)	11.00 (4.24)	10.00 (3.75)
Inhibition Switching Remote	10.50 (3.50)	10.50 (1.75)	10.50 (2.00)

Note. Mdn = median; IQR = interquartile range. In Person and Remote refer to administration condition; scores are: verbal fluency totals, RBANS index scores, z-scores for Oral Trail Making Test A and B, raw forward and backward Digit Span scores, and DKEFS

Colour Word Interference scaled scores. The Language Index shows a larger In-person/Remote difference within the younger age group (108.00 vs 96.00) than the older group (97.50 vs 97.00); this pattern is explored further below.

Primary Analysis: Consistency Between Administration Conditions

The primary research question concerned the reliability and comparability of neuropsychological test scores obtained from in-person and remote assessments. This was addressed using the Intraclass Correlation Coefficient (ICC), with Spearman's rank-order correlations calculated as a non-parametric check, given several variables showed deviations from normality.

Intraclass Correlation Coefficients

A two-way mixed-effects model with consistency was used to calculate single-measures ICCs for each primary cognitive variable, comparing in-person and remote scores across the whole sample ($N = 16$). Results are presented in Table 6.

Consistency between conditions ranged from poor to excellent across the primary measures. Excellent consistency was observed for the Oral Trail Making Test A (z-score) (ICC = .906), the RBANS Total Scale Score (ICC = .796), DKEFS Colour Naming (ICC = .808), and RBANS Attention Index (ICC = .766). Good consistency was found for several measures including FAS (ICC = .720), RBANS Visuospatial Index (ICC = .689), Digit Span Backward (ICC = .740), and DKEFS Word Reading (ICC = .671). All such ICCs were statistically significant (see Table 6). Fair consistency was observed for RBANS Immediate Memory Index (ICC = .471), RBANS Delayed Memory Index (ICC = .554), Oral TMT-B (z) (ICC = .469), Digit Span Forward (ICC = .412), DKEFS Inhibition (ICC = .512) and DKEFS Inhibition Switching (ICC = .515). Poor to fair, non-significant consistency was observed for the RBANS Language Index (ICC = .081, $p = .378$) and Animal fluency total (ICC = .403, $p = .054$), suggesting these measures showed less comparability between in-person and remote administration in this sample.

Table 6 *Intraclass correlation coefficients (single measures) between in-person and remote administration*

Measure	ICC	df	95% CI	p	Descriptor
Verbal Fluency FAS	.720	15, 15	[.364, .892]	< .001	Good

Verbal Fluency Animals	.403	15, 15	[−.099, .741]	.054	Fair
RBANS Immediate Memory Index	.471	15, 15	[−.014, .777]	.028	Fair
RBANS Visuospatial Index	.689	15, 15	[.311, .879]	.001	Good
RBANS Language Index	.081	15, 15	[−.417, .542]	.378	Poor
RBANS Attention Index	.766	15, 15	[.450, .912]	< .001	Excellent
RBANS Delayed Memory Index	.554	15, 15	[.098, .818]	.010	Fair
RBANS Total Scale Score	.796	15, 15	[.510, .924]	< .001	Excellent
Oral TMT-A (z)	.906	15, 15	[.753, .966]	< .001	Excellent
Oral TMT-B (z)	.469	15, 15	[−.017, .776]	.029	Fair
Digit Span Forward	.412	15, 15	[−.087, .746]	.050	Fair
Digit Span Backward	.740	15, 15	[.401, .901]	< .001	Good
DKEFS Colour Naming	.808	15, 15	[.535, .929]	< .001	Excellent
DKEFS Word Reading	.671	15, 15	[.280, .871]	.002	Good
DKEFS Inhibition	.512	15, 15	[.039, .797]	.018	Fair
DKEFS Inhibition Switching	.515	15, 15	[.044, .799]	.017	Fair

Note. ICC = single measures intraclass correlation coefficient (two-way mixed effects, consistency). df = degrees of freedom (df1, df2). CI = confidence interval.

As the low ICC for the RBANS Language Index appeared at odds with the reliable scores typically reported for verbally mediated tasks such as picture naming and semantic fluency a number of checks and further analyses were conducted.[2] As we found no evidence of clerical error or a pattern tied to modality, form, or age group, we can be confident in the accuracy of the scores recorded, but not confident that the RBANS Language Index's poor consistency in this sample would apply in other samples.

Spearman's Rank-Order Correlations

As a non-parametric check of the ICC findings, Spearman's rank-order correlations were obtained between in-person and remote scores for each primary measure. The pattern of results reflected the ICC findings, supporting the consistency estimates above. The strongest correlations were observed for the RBANS Delayed Memory Index ($r_s(14) = .859$, $p < .001$), Digit Span Backward ($r_s(14) = .850$, $p < .001$), RBANS Total Scale Score ($r_s(14) = .799$, $p < .001$), DKEFS Colour Naming ($r_s(14) = .787$, $p < .001$), and Oral TMT-A ($r_s(14) = .791$, $p < .001$). Consistent with the ICC results, the weakest and only non-significant correlation was for the RBANS Language Index ($r_s(14) = .059$, $p = .829$). Notably, the RBANS Delayed Memory Index showed only a Fair ICC (.554) despite showing the strongest Spearman correlation ($r_s(14) = .859$) illustrating that a measure can have a strong Spearman's correlation while still having a weaker ICC.

---

[2] First, Picture Naming and Fluency scores were double-checked against original records, and no clerical errors were identified. It was noted, however, that four participants made errors in Picture Naming in Session 2, following perfect performance in Session 1: two participants (one aged 18-40 and bilingual, one aged 65+) lost points on Remote RBANS Form A; one participant (aged 18-40 and bilingual) lost points on In-Person RBANS Form B; and one participant (aged 18-40) lost points on In-Person RBANS Form A. There hence seemed to be no clear pattern of errors that could be attributed to administration condition (two Remote, two In-person), RBANS form (three Form A, one Form B), or age group (three 18-40, one 65+). Notably, both of the sample's two bilingual participants were among the four who lost points, raising the possibility that specific Picture Naming items may be less accessible for individuals from different linguistic backgrounds. However, this observation is based on a very small number of cases, does not really account for errors being restricted to Session 2, and should be interpreted cautiously.

To investigate further, ICCs were calculated separately for the two Language subtests contributing to the RBANS Language index (as it was possible that the unusual pattern of naming scores drove the overall language ICC). However, both Picture Naming (ICC = -.121, 95% CI [-.570, .383], p = .678) and Semantic Fluency (ICC = -.078, 95% CI [-.540, .420], p = .617) showed negative, non-significant consistency, suggesting both subtests contributed to the poor RBANS Language Index ICC.

The ceiling effect on Picture Naming (where a loss of one point corresponds to a drop to as low as the 25th percentile) means there was limited variation in scores overall, which may also have suppressed reliability coefficients particularly within this small sample.

Table 7 *Spearman's rank-order correlations between in-person and remote scores*

Measure	r_s	p
FAS	.703	.002
Animals	.580	.019
Immediate Memory Index	.420	.105
Visuospatial Index	.549	.028
Language Index	.059	.829
Attention Index	.675	.004
Delayed Memory Index	.859	< .001
Total Scale Score	.799	< .001
TMT-A (z)	.791	< .001
TMT-B (z)	.509	.044
Digit Span Forward	.381	.146
Digit Span Backward	.850	< .001
Colour Naming	.787	< .001
Word Reading	.671	.004
Inhibition	.498	.049
Inhibition Switching	.510	.044

Note. rs = Spearman's rank-order correlation coefficient. N = 16 and df = 14 for all correlations.

To explore potential practice effects, descriptive statistics (median, interquartile range, minimum, and maximum) were calculated for all primary cognitive variables separately for Session 1 and Session 2, irrespective of administration modality (i.e., in-person / remote). The results are presented in Appendix M, Practice Effects Descriptives. The pattern of scores

was mixed: some measures showed modest improvements from Session 1 to Session 2 (notably FAS verbal fluency, Mdn = 42.00 vs 49.50, and RBANS Visuospatial Index, Mdn = 101.00 vs 109.00), while others remained stable (e.g., RBANS Immediate Memory Index, Digit Span Forward, DKEFS Word Rading) or showed slight decreases (e.g., RBANS Attention Index, Oral TMT-A z-score, Digit Span Backward). No consistent pattern of practice effects was observed across all measures, consistent with the counterbalanced design and the use of parallel forms for the RBANS and alternate trial sets for Digit Span,

Secondary Analysis: Differences by Administration Condition, Age Group, and Testing Order

The secondary aim explored whether test performance differed significantly according to administration condition (in-person vs remote), age group (younger vs older), or testing order, across the primary cognitive domains. This was addressed using a repeated-measures ANOVA, with Condition specified as the within-subjects factor, Age Group and testing order as between-subjects factors, and years of education included as a covariate given that education differed between age groups (see Preliminary Analyses).

Effect of Administration Condition

The multivariate test for the overall effect of administration condition (combined across all 16 primary measures) was not statistically significant, Pillai's Trace = .831, $F(11, 1) = 0.45$, $p = .837$, $\eta_2p = .831$. At the univariate level, no individual measure showed a statistically significant main effect of condition (all ps > .10), consistent with the ICC and Spearman findings above. This is also consistent with the fact that no directional hypotheses were made about condition differences. Effect sizes for the condition effect were generally small (η_2p ranging from .002 for Digit Span Backward to .117 for the RBANS Immediate Memory Index), with the exception of Digit Span Forward, which showed a medium to large effect ($\eta_2p = .225$) despite non-significance, $F(1,11) = 3.19$, $p = .102$, indicating a possible difference between in-person and remote performance on this measure that the study may not have had sufficient power to detect.

Effects of Age Group and Testing Order

It was also considered important to explore potential interactions between administration format and age, and we tentatively hypothesised that older adults would show less consistency in performance between administration conditions than younger adults, given the potential impact of ageing, a wider expected range of cognitive abilities, and possible technological or sensory challenges specific to remote testing. This was tested via the

Condition x Age Group interaction in the ANOVA The interaction term was not statistically significant (Pillai's Trace = .963, F(11, 1) = 2.34, p = .473, η^2p = .963). At the univariate level, no individual measure showed a statistically significant interaction between condition and age group (all ps > .33), with effect sizes ranging from minimal (η^2p = .002 for RBANS Total Scale Score) to small-to-moderate (η^2p = .085 for the RBANS Visuospatial and Language Indexes (see Appendix N) for the full breakdown across all 16 measures. This provides no evidence to support the hypothesis that the effect of administration condition on performance differed between age groups.

No primary measure showed a statistically significant main effect of age group (all ps > .10). A significant main effect of testing order (S1_Condition) was found for the Delayed Memory Index (F(1, 11) = 7.06, p = .022, η^2p = .391 (a large effect) albeit not corrected for multiple comparisons), with effects for the RBANS Total Scale Score (F(1, 11) = 4.78, p = .051, η^2p = .303), and Immediate Memory Index (F(1, 11) = 4.01, p = .071 η^2p = .267 (both large effects) approaching significance thresholds, suggesting some sensitivity to the order in which testing modalities were completed for memory-based measures. The descriptive practice-effects data (see Appendix M) do not offer a clear explanation for this: the increase in median scores between Session 1 and Session 2 was not marked for any of these measures (RBANS Delayed Memory Index: Mdn = 100.00 to 101.00; RBANS Total Scale Score: Mdn = 99.50 to 101.00; RBANS Immediate Memory Index: Mdn = 103.00 at both sessions). This suggests the testing order effect is unlikely to be driven by straightforward practice effect, though this should be interpreted cautiously given the small sample.

Age Group by Testing Order Interaction

Significant interactions between age group and testing order were found for the following DKEFS Colour Word Interference subtests: Colour Naming, F(1, 11) = 9.66, p = .010, η^2p = .468; Inhibition, F(1, 11) = 14.10, p = .003, η^2p = .562; and Word Reading, F(1, 11) = 5.02, p = .047, η^2p = .313, all large effects. A further interaction was near the threshold for significance for the Attention Index, F(1, 11) = 3.63, p = .083, η^2p = .248 (also a large effect). These findings suggest that the effect of testing order on performance may differ between younger and older adults for measures of processing speed, inhibition, and attention, though this should be interpreted cautiously given the small number of participants in each age group and order combination (n=4).

Estimated marginal means (controlling for Education Years, see Table 8), revealed a consistent crossover pattern across all four measures: younger participants scored somewhat higher when they had completed in-person testing first (versus remote first), whereas older participants showed the opposite pattern, scoring higher when they had completed remote testing first (versus in-person first). This reversal was most obvious for the RBANS Attention Index, where older participants who tested remotely first scored, on average, 13.7 points higher than those who tested in-person first, compared to a minimal 1.1-point difference in the younger group.

Table 8 *Estimated marginal means (controlling for Education Years) for Age Group and testing order, for measures showing a significant or near significant interaction*

Measure	Younger In-person-first	Younger Remote-first	Older In-person-first	Older Remote-first
Colour Naming	9.67	7.89	5.94	10.63
Word Reading	10.76	8.89	8.99	11.37
Inhibition	12.25	10.09	9.12	12.17
Attention Index	99.94	98.80	96.71	110.43

Note. Means represent scores averaged across in-person and remote administration, controlling for participants' years of education. In-person first and Remote first refer to which administration format participants completed first (in Session 1).

Effect of Education

Education, included as a covariate given the group difference identified in the Preliminary Analyses, was significantly associated with performance on the RBANS Attention Index, $F(1, 11) = 7.85$, $p = .017$, $\eta_2p = .417$; Digit Span Forward, $F(1, 11) = 11.37$, $p = .006$, $\eta_2p = .508$; and Digit Span Backward, $F(1, 11) = 6.11$, $p = .031$, $\eta_2p = .357$, all large effects, indicating that years of education accounted for a meaningful proportion of variance on these measures independent of condition, age group, or testing order.

Summary

In line with the primary research question, scores obtained from remote and in-person administration were generally consistent for the majority of primary cognitive measures. Oral Trail Making Test A, RBANS Total Scale Score, DKEFS Colour Naming, and RBANS Attention Index showed the strongest agreement between conditions. Consistency was more

variable for other measures: the RBANS Language Index showed poor, non-significant consistency, while Immediate Memory (ICC = .471), Delayed Memory (ICC = .554), Animal Fluency, Oral TMT-B (z), Digit Span Forward, and the DKEFS Inhibition subtests showed fair consistency. This suggests that, in this sample, these measures showed lower consistency between formats in this sample. Given the limited sample size, these findings should not be interpreted as establishing that the measures are less suitable for remote administration.

The poor consistency of the RBANS Language Index ICC was evident across both component subtests, Picture Naming and Semantic Fluency, and was not attributable to a clerical error, administration condition, or RBANS form. However, this finding should be interpreted cautiously given the small sample size.

There was no significant overall effect of administration condition on performance, nor evidence that the effect of administration condition differed between younger and older adults. Testing order was associated with performance on some memory measures. For Colour Naming, Word Reading, and Inhibition, younger participants tended to perform better when remote testing was completed first. The RBANS Attention Index showed a similar, near-significant pattern. Given the limited power and the small number of participants in each age-group and testing-order combination, these findings require replication. Their interpretation in relation to the wider teleneuropsychology literature is considered in the Discussion.

Discussion

Summary of Results

The current study examined whether neuropsychological cognitive test scores obtained remotely were comparable with those obtained face-to-face in younger and older adults. Overall, most measures showed broadly consistent scores across administration formats, and there was no evidence of a general effect of administration format on performance or that this effect differed by age group.

The findings nevertheless identified measure specific variation. The RBANS Language Index showed poor consistency between formats in this sample; however, this isolated finding requires replication and may reflect chance variation. An exploratory age-related pattern was also observed in the effects of testing order on selected executive function measures. These findings are considered below in relation to existing teleneuropsychology literature, the study's methodological strengths and limitations, and implications for clinical neuropsychology.

Relation to Existing Findings from the Literature

Overall Reliability of Remote Administration

The level of consistency found between face-to-face and remote scores generally supports previous findings. Galusha-Glasscock et al. (2016) found RBANS correlations between face-to-face and video teleconference conditions ranging from r = .59 to .90 in a sample of 18 adults aged 58-84 with and without cognitive impairment. The present study's *excellent* range ICCs for the RBANS Total Scale Score and Attention Index are well within that range. This pattern is also broadly in line with professional guidance as the British Psychological Society (2020) suggests that straightforward, verbally delivered tasks translate most easily to remote formats. Further, Brearly and colleagues' (2017) meta-analysis similarly found digit span, verbal fluency, and word list learning tasks to be largely unaffected by videoconference delivery. FAS fluency (ICC = .720) and Digit Span Backward (ICC = .740) both reached the *Good* range, also supporting this pattern. However, not every measure followed this pattern of previous findings as Animal fluency and more surprisingly, the RBANS Language Index performed comparatively poorly despite also being verbal tasks, a difference further explored below.

The RBANS Language Index and Picture Naming

The poor consistency of the RBANS Language Index was unexpected, as it did not fit with the broader pattern of results in the present study or with the findings of Galusha-Glasscock et al. (2016), who did not report a comparable problem with this index. Given the small sample, the number of measures analysed, and the fact that this result was not expected based on previous research, the poor Language Index finding may represent chance variation. It should therefore not be interpreted as evidence that remote administration compromises the reliability of the Language Index.

Exploratory review of the data identified score variation in Picture Naming and Semantic Fluency, including among the two bilingual participants. However, this pattern was not clearly associated with administration condition, RBANS form, or age group, and does not explain the observed inconsistency. The Picture Naming subtest has a narrow scoring range of 10 items, so a difference of one point may produce a large change in the index score. Differences in cultural or linguistic familiarity with individual items may also be relevant. However, these observations are post hoc and are only possible explanations. They do not show that remote administration, participants' language background or particular test items caused the poor consistency. Replication in a larger and more linguistically diverse sample is needed before drawing conclusions about remote administration of the RBANS Language Index.

Observations and Participant Feedback

Informal feedback received from participants after testing sessions offers some tentative, qualitative context for the quantitative findings, though these observations were not part of the study's formal data collection. Among younger participants (18-40), preference for administration format appeared to match with participants' everyday working arrangements: those familiar with remote working reported feeling more comfortable testing from home, whereas those who are used to working in person reported a preference for face-to-face testing. One participant described feeling distracted at home during verbal fluency tasks, reporting that they relied on prompts from their immediate environment rather than generating items imaginatively, as they had done during the face-to-face session, and felt this limited their performance remotely. This observation offers a plausible though speculative account of how environmental context might influence performance on generative tasks. Participants also occasionally described feeling fatigued at certain sessions, for example

following a night shift or poor sleep; although appointments were scheduled as flexibly as possible and participants were advised to ensure they were well rested, scheduling was ultimately constrained by the availability of testing space and researcher and participant availability, meaning this could not always be fully controlled.

Among the older age group (65 and over), the majority expressed a preference for face-to-face testing, with most attributing this preference to being retired and unfamiliar with remote formats. Several described face-to-face testing as feeling more like 'real' test conditions, consistent with their prior experience of being assessed. Interestingly, subjective preference did not always align with objective performance, illustrating that self-reported ease of testing may not necessarily correspond to actual performance.

Verbal Fluency and Practice Effects

In contrast to the Language Index, Animal fluency showed comparatively higher, though still non-significant, consistency between conditions (ICC = .403). Scores also increased modestly from Session 1 to Session 2, suggesting a practice effect. FAS fluency showed a similar pattern, increasing from a median of 42.00 in Session 1 to 49.50 in Session 2, potentially reflecting reduced task anxiety or an increase in familiarity with the task on the second session, consistent with previous findings for verbal fluency tasks (Bartels et al., 2010; Collie et al., 2003). This pattern was not observed consistently across measures: the RBANS Visuospatial Index also increased across sessions, several measures (RBANS Immediate Memory Index, Digit Span Forward, DKEFS Word Reading) showed minimal change, and a small number of measures (RBANS Attention Index, Oral TMT-A, Digit Span Backward) decreased slightly. This variability is consistent with the study's design, which incorporated parallel RBANS forms and alternate Digit Span trial sets specifically to reduce practice effects for memory-based measures. Overall, verbal fluency performance was consistent with expectations from existing literature, supporting the interpretation that the Language Index's poor reliability is attributable to Picture Naming rather than Semantic Fluency. The environmental cue observation described above offers one plausible, exploratory explanation for why verbal fluency performance may vary between formats beyond practice effects alone, though this remains speculative.

Testing Order and Age

Among the more unexpected results was a significant interaction between testing order and age group across several DKEFS measures of executive function, inhibition and processing speed, i.e., Colour Naming, Inhibition, and Word Reading were all significant, with RBANS Attention Index showing a near significant effect. Participants in the younger age group tended to do slightly better having completed the face-to-face session first, while the older age group participants showed the reverse pattern, performing better when the remote session was first. The effect was largest for the Attention Index, where older participants who completed remote first scored on average 13.7 points higher than those who completed face-to-face testing first. Among the younger age group, this same comparison showed a much smaller difference of just 1.1 points.

Research has tended to focus on whether older adults do worse with remote testing generally, rather than whether the order of testing interacts with age specifically. It may be that starting with remote testing was simply less stressful for older participants which then carried over into the second session. Whereas the younger age group who are generally more at ease and familiar with video call technology would have less gain from this. Testing order was built into the design mainly as a counterbalancing tool rather than to test a specific age-related hypothesis and therefore this result is more of an interesting finding than a confirmed effect and is a finding for replication rather than one to build conclusions on.

Testing order also affected memory performance independently of age, with a significant effect on the RBANS Delayed Memory Index and near significant effects on the Total Scale Score and Immediate Memory Index. Session 1 and Session 2 medians did not differ substantially for these measures, suggesting the effect is not attributable to practice. The mechanism underlying this testing order effect therefore remains unclear and similar to the interaction reported above, warrants further exploration.

Strengths and Limitations of the Study

Several design choices strengthened confidence in these findings. For example, administration condition, RBANS form, and testing order were all counterbalanced, reducing the risk that findings were confounded by testing sequence. Reliability was assessed using several complementary methods (ICC, Spearman's rank order correlations, and ANOVA), and effect sizes provided a fuller picture than p-values alone given the study's modest power.

Limitations are equally worth highlighting. For example, sample size has been flagged as a limitation with just 16 participants, the study had 80% power to detect only fairly substantial

correlations (r = .636) and large MANOVA effects (f^2 = .738), and therefore more modest effects could easily have gone undetected, and any results drawn from a small subgroup (e.g., the n = 4 Age Group by testing order) should be interpreted with caution. No power calculation was carried out at the level of the four-way counterbalancing, though this was not a realistic calculation to attempt given the overall sample, but it is acknowledged as a limitation.

No correction for multiple comparisons was applied across the 16 ICCs, 16 Spearman's correlations, and multiple ANOVA effects tested, consequently, results close to the conventional significance threshold (e.g., the testing order effect on the RBANS Total Scale Score, p = .051) may in some cases reflect chance variation rather than a genuine effect. The breadth of the cognitive battery also represented a limitation in relation to the available sample size and resources. Although including multiple measures allowed an exploratory assessment of several cognitive domains, it resulted in a large number of analyses relative to the sample and limited the statistical power available for each measure. A more focused battery, with one pre-specified primary measure for each key domain, for example, memory, executive function, language and processing speed would have reduced the number of comparisons and allowed clearer, more adequately powered conclusions.

A further limitation to acknowledge is that performance validity was not formally examined using a standalone performance validity test or an embedded validity indicator. To minimise participant burden and fit the assessment within a 60-minute appointment, the battery prioritised primary cognitive measures. Concerns about suboptimal test engagement were not anticipated in this voluntary research context. However, the RBANS includes an embedded Effort Index, which combines performance across two subtests to identify potentially insufficient engagement (Silverberg et al., 2007). This index was not examined because the analyses prioritised the study's primary cognitive measures rather than additional embedded performance validity indicators. Future studies should consider including standalone performance validity tests and / or examining validated embedded indicators, particularly when assessing clinical populations or other contexts in which the validity of test performance may be less clear.

The two-to-four-week gap between sessions, chosen to reduce the risk of practice effects is longer than the same day interval used by Galusha-Glasscock et al. (2016), and may have introduced natural variation in participants' performance between sessions, for example due

to differences in mood, fatigue, or alertness on the day of testing, independent of testing modality. The scheduling constraints described in participant feedback above whereby testing depended on the availability of both participants and researcher, as well as room booking, could not always guarantee participants were well rested however this reflects a real limitation in clinical practice. This could have affected reliability estimates for measures particularly sensitive to this kind of variation. Additionally, participants were healthy adults recruited in Scotland, which limits how confidently these results can be generalised to other populations, regions, or linguistic and cultural contexts.

Theoretical Implications

Overall, these findings support the position that remote neuropsychological assessment can produce results comparable to face-to-face testing across a wide range of cognitive domains, consistent with existing literature and professional guidance. However, the unexpected finding for the Language Index warrants further investigation. Given that it did not fit with the wider pattern of results, the small sample, and the number of analyses undertaken, it may reflect chance variation rather than a reliable limitation of remote administration. These findings raise an important point for reliability research conducted with small samples. Although clerical error and obvious associations with administration condition, form, or age group were not identified, this does not establish that the finding reflects a reliable pattern for the Language Index. With only four participants contributing to this pattern, replication in a larger sample is needed before drawing conclusions about the subtest more broadly.

The testing order by age group interaction represents a further consideration for future research. The order in which assessment formats are administered may affect younger and older adults differently, an area that has received limited attention in literature which tends to primarily focus on whether older adults are generally disadvantaged by remote testing. Participant feedback offers some real-world context to this point, i.e., older participants' general unfamiliarity with remote formats, and the mismatch observed between participant subjective preference and their objective performance, suggests that self-reported comfort with a testing format may not be straight forward in predicting how that format affects performance. Together these findings support the broader use of remote neuropsychological assessment, while highlighting practical considerations. However, the present findings were obtained from healthy volunteers and should not be generalised directly to clinical populations with cognitive impairment or disability. People referred for neurological assessment may experience difficulties that were not represented in this sample, including

more substantial cognitive impairment, fatigue, sensory or motor difficulties, communication needs, reduced familiarity with technology, and a need for support to access or complete remote assessment. These factors may affect both the feasibility of remote testing and the comparability of scores obtained remotely and face-to-face. Remote assessment should therefore be considered on an individual basis and alongside clinical judgement, rather than on the basis of the present findings alone. Further research is needed in clinically representative populations before these measures can be considered equivalent across administration formats in routine practice. Within these limitations, the findings highlight practical considerations for future research and clinical implementation. Firstly, the poor consistency observed for the RBANS Language Index in this sample requires replication before any conclusions can be drawn about its use across remote and face-to-face assessment. Secondly, the interaction between testing order and age group suggests that assessment sequence may affect performance differently across age groups and should be considered when interpreting repeated remote assessments in mixed age clinical populations.

Future studies comparing remote and face-to-face neurological assessment with relatively small samples would benefit from selecting a smaller number of pre-specified measures across key cognitive domains. This would allow sample size calculations to be based on clearly defined primary outcomes, reduce the number of statistical comparisons and provide more precise estimates of consistency between formats. Larger, adequately powered studies could use a wider battery of measures and examine whether findings generalise to clinical populations.

References

Anderson, M., & Perrin, A. (2017). Tech adoption climbs among older adults. Pew Research Center. https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/

Barney, A., Buckelew, S., Mesheriakova, V., & Raymond-Flesch, M. (2020). The COVID-19 pandemic and rapid implementation of adolescent and young adult telemedicine: Challenges and opportunities for innovation. *Journal of Adolescent Health, 67*(2), 164-171. https://doi.org/10.1016/j.jadohealth.2020.05.006

Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: A longitudinal study on frequent repetitive cognitive testing. BMC Neuroscience, 11, 118. https://doi.org/10.1186/1471-2202-11-118

Belleville, S., Mellah, S., Boller, B., & Ouellet, É. (2023). Activation changes induced by cognitive training are consistent with improved cognitive reserve in older adults with subjective cognitive decline. *Neurobiology of Aging, 121*, 107-118. https://doi.org/10.1016/j.neurobiolaging.2022.10.010

Bergman, I., Franke Föyen, L., Gustavsson, A., & Van den Hurk, W. (2025). Test–retest reliability, practice effects and estimates of change: A study on the Mindmore digital cognitive assessment tool. Scandinavian Journal of Psychology, 66(1), 1-14. https://doi.org/10.1111/sjop.13054

Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspy, S. R., Harder, L., Kanter, G., Lanca, M., Lechuga, D. M., Morgan, J. M., Most, R., Puente, A. E., Salinas, C. M., & Woodhouse, J. (2020). Inter organizational practice committee recommendations/guidance for teleneuropsychology in response to the COVID-19 pandemic. *Archives of Clinical Neuropsychology, 35*(6), 647-659. https://doi.org/10.1093/arclin/acaa046

British Psychological Society. (2021). BPS code of human research ethics (3rd ed.). British Psychological Society.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6(4), 284-290. https://doi.org/10.1037/1040-3590.6.4.284

Collie, A., Maruff, P., Darby, D. G., & McStephen, M. (2003). The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. Journal of the International Neuropsychological Society, 9(3), 419-428.

Cullum, C. M., Hynan, L. S., Grosch, M., Parikh, M., & Weiner, M. F. (2014). Teleneuropsychology: Evidence for video teleconference-based neuropsychological assessment. Journal of the International Neuropsychological Society, 20(10), 1028-1033. https://doi.org/10.1017/S1355617714000873

Cyr, A.-A., Romero, K., & Galin-Corini, L. (2021). Online cognitive testing of older adults: In-person versus at-home: Do we get the same results? *JMIR Aging, 4*(2), e23384. https://doi.org/10.2196/23384

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). Delis-Kaplan Executive Function System (D-KEFS). Psychological Corporation.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.* Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Fox-Fuller, J. T., Ngo, J., Pluim, C. F., Kaplan, R. I., Kim, D.-H., Anzai, J. A. U., Yucebas, D., Briggs, S., Aduen, P., Cronin-Golomb, A., & Quiroz, Y. T. (2021). Initial investigation of test–retest reliability of home-to-home teleneuropsychological assessment in healthy, English-speaking adults. *The Clinical Neuropsychologist, 36*(8), 2153-2167. https://doi.org/10.1080/13854046.2021.1954244

Galusha-Glasscock, J. M., Horton, D. K., Weiner, M. F., & Cullum, C. M. (2016). Video teleconference administration of the Repeatable Battery for the Assessment of Neuropsychological Status. Archives of Clinical Neuropsychology, 31(1), 8-11. https://doi.org/10.1093/arclin/acv058

Gardner, M., Aslanzadeh, F., Zarrella, G., Braun, S., Loughan, A., & Parsons, M. (2021). Cancer, cognition, and COVID: Delivering direct-to-home teleneuropsychology services to neuro-oncology patients. Neuro-Oncology Practice, 8(4), 485-496. https://doi.org/10.1093/nop/npab021

Jaffe, D. H., Lee, L., Huynh, S., & Haskell, T. P. (2020). Health inequalities in the use of telehealth in the United States in the lens of COVID-19. Population Health Management, 23(5), 368-377. https://doi.org/10.1089/pop.2020.0186

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford University Press.

Lezak, M. D. (Ed.). (1995). Neuropsychological assessment (3rd ed.). Oxford University Press.

Mrazik, M., Millis, S., & Drane, D. L. (2010). The Oral Trail Making Test: Effects of age and concurrent validity. Archives of Clinical Neuropsychology, 25(3), 236-243. https://doi.org/10.1093/arclin/acq006

Parikh, M., Grosch, M. C., Graham, L. L., Hynan, L. S., Weiner, M. F., Shore, J. H., & Cullum, C. M. (2013). Consumer acceptability of brief videoconference-based neuropsychological assessment in older individuals with and without cognitive impairment. *The Clinical Neuropsychologist, 27*(5), 808-817. https://doi.org/10.1080/13854046.2013.791725

Parsons, M. W., Gardner, M. M., Sherman, J. C., Pasquariello, K., Grieco, J. A., Kay, C. D., Pollak, L. E., Morgan, A. K., Carlson-Emerton, B., Seligsohn, K., Davidsdottir, S., Pulsifer, M. B., Zarrella, G. V., Burstein, S. M., & Mancuso, S. M. (2022). Feasibility and acceptance of direct-to-home tele-neuropsychology services during the COVID-19 pandemic. *Journal of the International Neuropsychological Society, 28*(2), 210-215. https://doi.org/10.1017/S1355617721000436

Parsons, T. D. (2016). Future prospects for a computational neuropsychology. In T. D. Parsons (Ed.), Clinical neuropsychology and technology: What's new and how we can use it (pp. 135-146). Springer. https://doi.org/10.1007/978-3-319-31075-6_8

Pearson. (2016). *Q-global* [Web-based platform].

Pearson. (2020, March 23). *UK telepractice no objection letter*. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/telepractice/telepractice-no-objection-letter.pdf

Randolph, C. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status. The Psychological Corporation.

Schmand, B. (2019). Why are neuropsychologists so reluctant to embrace modern assessment techniques? The Clinical Neuropsychologist, 33(2), 209-219. https://doi.org/10.1080/13854046.2018.1523468

Silverberg, N. D., Wertheimer, J. C., & Fichtenberg, N. L. (2007). An effort index for the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). *The Clinical Neuropsychologist, 21*(5), 841-854. https://doi.org/10.1080/13854040600850958

Singh, S., & Germine, L. (2021). Technology meets tradition: A hybrid model for implementing digital tools in neuropsychology. International Review of Psychiatry, 33(4), 382-393.

Sperling, S. A., Acheson, S. K., Fox-Fuller, J., Colvin, M. K., Harder, L., Cullum, C. M., Randolph, J. J., Carter, K. R., Espe-Pfeifer, P., Lacritz, L. H., Arnett, P. A., & Gillaspy, S. R. (2023). Tele-neuropsychology: From science to policy to practice. Archives of Clinical Neuropsychology, 39(2), 227-248. https://doi.org/10.1093/arclin/acad066

Sumpter, R., Camsey, E., Meldrum, S., Alford, M., Campbell, I., Bois, C., O'Connell, S., & Flood, J. (2022). Remote neuropsychological assessment: Acceptability and feasibility of direct-to-home teleneuropsychology methodology during the COVID-19 pandemic. *The Clinical Neuropsychologist, 37*(2), 432-447. https://doi.org/10.1080/13854046.2022.2056922

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. Archives of Clinical Neuropsychology, 14(2), 167-177.

Urbaniak, G. C., & Plous, S. (2019). Research Randomizer [Computer software]. https://www.randomizer.org/

Van Patten, R. (2021). Introduction to the special issue—Neuropsychology from a distance: Psychometric properties and clinical utility of remote neurocognitive tests. Journal of Clinical and Experimental Neuropsychology, 43(8), 767-773.

Wadsworth, H., Dhima, K., Womack, K., Hart, J., Weiner, M., Hynan, L., & Cullum, M. (2018). Validity of teleneuropsychological assessment in older patients with cognitive

disorders. Archives of Clinical Neuropsychology, 33(8), 1040-1045. https://doi.org/10.1093/arclin/acx140

Appendices

Appendix A. PRISMA reporting guidelines

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	P7
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	P8
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	P9
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	P11-12
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	p13
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	P15
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	p15
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	p16
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	P16
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	P16-20
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	P16-18
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	p16
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	P19
Synthesis	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and	P20

Section and Topic	Item #	Checklist item	Location where item is reported
methods		comparing against the planned groups for each synthesis (item #5)).	
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	p16
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	P19
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	p24
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	P17
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	P17
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	n/a
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	p81
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	P19
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	P19
Study characteristics	17	Cite each included study and present its characteristics.	p20
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	p81
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	p22-24
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	p24
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	p22-24
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	p22-24
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	p81
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	n/a
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	p81
DISCUSSION			

Section and Topic	Item #	Checklist item	Location where item is reported
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	P26
	23b	Discuss any limitations of the evidence included in the review.	P28
	23c	Discuss any limitations of the review processes used.	P28
	23d	Discuss implications of the results for practice, policy, and future research.	P29
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	P8
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Not provided
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	p13
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	P8
Competing interests	26	Declare any competing interests of review authors.	n/a
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Not provided

Appendix B. Search Strategies

Database MEDLINE (ovid)		Search Strategy
	1.	("brain injur*" or "acquired brain injur*" or "traumatic brain injur*" or TBI or "head injur*" or "head trauma*" or stroke* or "hypoxic brain" or "anoxic brain" or "hypoxic-ischemic encephalopathy" or encephalitis or "cerebr* injur*" or "cerebr* trauma*").ti,ab. or exp Brain Injuries/ or exp Craniocerebral Trauma/ or exp Stroke/ or exp Brain Ischemia/ or exp Hypoxia, Brain/ or exp Encephalitis/
	2.	(telehealth or telemedicine or telerehab* or "tele-rehabilitation" or telepsycholog* or teletherap* or eHealth or "e-health" or mHealth or "web-based" or "internet-based" or online or videoconferenc* or "video call*" or telephone or phone or smartphone* or app or apps or digital).ti,ab. or exp Telemedicine/ or exp Remote Consultation/ or exp Internet/ or exp Videoconferencing/
	3.	("cognitive behavio* therap*" or CBT or "cognitive therap*" or "cognitive restructuring" or "behavio* therap*" or "behavio* intervention*" or "problem solving therap*" or "problem-solving therap*" or PST).ti,ab. or exp Cognitive Behavioral Therapy/ or exp Behavior Therapy/
	4.	1 and 2 and 3
	5.	randomized controlled trial.pt. or controlled clinical trial.pt. or randomi#ed.ti,ab. or placebo.ti,ab. or "clinical trial*".ti,ab. or trial*.ti,ab. or feasibility.ti,ab. or pilot.ti,ab. or acceptability.ti,ab. or cohort.ti,ab. or observational.ti,ab. or non-randomi#ed.ti,ab. or nonrandomi#ed.ti,ab. or "pre and post*".ti,ab. or "pre-post*".ti,ab. or "service evaluation*".ti,ab. or stud*.ti,ab.
	6.	4 and 5
	7.	exp animals/ not humans.sh.
	8.	6 not 7

Database Embase (ovid)		Search Strategy
	1.	exp brain injury/ or exp cerebrovascular accident/ or exp brain hypoxia/ or exp brain ischemia/ or exp encephalitis/
	2.	(brain injur* or acquired brain injur* or traumatic brain injur* or TBI or head injur* or head trauma* or stroke* or hypoxic brain or anoxic brain or encephalitis).ti,ab.
	3.	1 or 2
	4.	exp telemedicine/ or exp internet/ or exp videoconferencing/
	5.	(telehealth or telemedicine or telerehab* or teletherap* or ehealth or mhealth or web-based or internet-based or online or videoconferenc* or video call* or telephone or phone or smartphone* or app or digital).ti,ab.
	6.	4 or 5
	7.	exp cognitive behavioral therapy/ or exp cognitive therapy/ or exp behavior therapy/
	8.	(cognitive behavio* therap* or CBT or cognitive therap* or cognitive restructuring or behavio* therap* or behavio* intervention* or problem solving therap* or problem-solving therap* or PST).ti,ab.
	9.	7 or 8

	10. 3 and 6 and 9
	11. random*.ti,ab. or exp crossover procedure/ or exp double blind procedure/ or exp single blind procedure/ or placebo.ti,ab. or trial*.ti,ab. or feasibility.ti,ab. or pilot.ti,ab. or acceptability.ti,ab. or cohort.ti,ab. or observational.ti,ab. or non-randomi#ed.ti,ab. or nonrandomi#ed.ti,ab. or "pre and post*".ti,ab. or "pre-post*".ti,ab. or "service evaluation*".ti,ab. or stud*.ti,ab.
	12. 10 and 11
	13. exp animal/ or exp invertebrate/ or animal experiment/ or nonhuman/
	14. exp human/
	15. 13 and 14
	16. 13 not 15
	17. 12 not 16

Database	**Search Strategy**
Cochrane CENTRAL	
	1. MeSH descriptor: [Brain Injuries]
	2. MeSH descriptor: [Stroke]
	3. MeSH descriptor: [Hypoxia-Ischemia, Brain]
	4. MeSH descriptor: [Encephalitis]
	5. ((brain NEXT injur*) or (acquired NEXT brain NEXT injur*) or (traumatic NEXT brain NEXT injur*) or TBI or (head NEXT injur*) or (head NEXT trauma*) or stroke* or "hypoxic brain" or "anoxic brain" or encephalitis or (cerebr* NEXT injur*) or (cerebr* NEXT trauma*)):ti,ab,kw
	6. #1 or #2 or #3 or #4 or #5
	7. (telehealth or telemedicine or telerehab* or teletherap* or eHealth or mHealth or "web-based" or "internet-based" or online or videoconferenc* or (video NEXT call*) or telephone or phone or smartphone* or app or digital):ti,ab,kw
	8. MeSH descriptor: [Telemedicine]
	9. #7 or #8
	10. MeSH descriptor: [Cognitive Behavioral Therapy]
	11. ((cognitive NEXT behavio* NEXT therap*) or CBT or (cognitive NEXT therap*) or "cognitive restructuring" or (behavio* NEXT therap*) or (behavio* NEXT intervention*) or (problem NEXT solving NEXT therap*) or (problem-solving NEXT therap*) or PST):ti,ab,kw
	12. #10 or #11
	13. (random* or trial* or "controlled trial" or feasibility or pilot or acceptability or cohort or observational or non-randomi#ed or nonrandomi#ed or (pre NEXT post*) or "pre-post*" or (service NEXT evaluation*) or stud*):ti,ab,kw
	14. #6 and #9 and #12 and #13

Database	**Search Strategy**
APA PsycINFO (EBSCO)	
	1. (DE "Traumatic Brain Injury" OR DE "Brain Concussion" OR DE "Chronic Traumatic Encephalopathy" OR DE "Post-Concussive Symptoms") OR (DE "Cerebrovascular Accidents" OR DE "Cerebral Ischemia" OR DE "Ischemic Stroke" OR DE "Cerebral Infarction") OR (DE "Head Injuries" OR DE "Brain

	Injuries" OR DE "Periventricular Leukomalacia" OR DE "Traumatic Brain Injury" OR DE "Brain Concussion" OR DE "Post-Concussive Symptoms" OR DE "Chronic Traumatic Encephalopathy") OR DE "Anoxia" OR (DE "Encephalitis" OR DE "Encephalomyelitis")
2.	TI ("brain injur*" OR "acquired brain injur*" OR "traumatic brain injur*" OR TBI OR "head injur*" OR "head trauma*" OR stroke* OR "hypoxic brain" OR "anoxic brain" OR encephalitis) OR AB ("brain injur*" OR "acquired brain injur*" OR "traumatic brain injur*" OR TBI OR "head injur*" OR "head trauma*" OR stroke* OR "hypoxic brain" OR "anoxic brain" OR encephalitis)
3.	S1 OR S2
4.	(DE "Telemedicine" OR DE "Teleconferencing" OR DE "Teleconsultation" OR DE "Telepsychiatry" OR DE "Telepsychology" OR DE "Telerehabilitation" OR DE "Videoconferencing" OR DE "Online Therapy") OR (DE "Internet" OR DE "Internet Access" OR DE "Internet Usage") OR DE "Videoconferencing"
5.	TI (telehealth OR telemedicine OR telerehab* OR teletherap* OR eHealth OR mHealth OR "web-based" OR "internet-based" OR online OR videoconferenc* OR "video call*" OR telephone OR phone OR smartphone* OR app OR digital) OR AB (telehealth OR telemedicine OR telerehab* OR teletherap* OR eHealth OR mHealth OR "web-based" OR "internet-based" OR online OR videoconferenc* OR "video call*" OR telephone OR phone OR smartphone* OR app OR digital)
6.	S4 OR S5
7.	DE "Cognitive Therapy" OR DE "Cognitive Behavior Therapy" OR DE "Cognitive Restructuring" OR DE "Cognitive Techniques" OR DE "Behavior Therapy" OR DE "Trauma-Focused Cognitive Behavior Therapy"
8.	TI ("cognitive behavio* therap*" OR CBT OR "cognitive therap*" OR "cognitive restructuring" OR "behavio* therap*" OR "behavio* intervention*" OR "problem solving therap*" OR "problem-solving therap*" OR PST) OR AB ("cognitive behavio* therap*" OR CBT OR "cognitive therap*" OR "cognitive restructuring" OR "behavio* therap*" OR "behavio* intervention*" OR "problem solving therap*" OR "problem-solving therap*" OR PST)
9.	S7 OR S8
10.	S3 AND S6 AND S9
11.	TI (random* OR trial* OR "controlled trial" OR feasibility OR pilot OR acceptability OR cohort OR observational OR non-randomi#ed OR nonrandomi#ed OR "pre and post*" OR "pre-post*" OR "service evaluation*" OR stud*) OR AB (random* OR trial* OR "controlled trial" OR feasibility OR pilot OR acceptability OR cohort OR observational OR non-randomi#ed OR nonrandomi#ed OR "pre and post*" OR "pre-post*" OR "service evaluation*" OR stud*) OR DE "Treatment Effectiveness Evaluation" OR DE "Clinical Trials"
12.	S10 AND S11
	Limiters applied to final line: Population Group - Human.

Appendix C. CCAT Quality Appraisal Ratings for all six included RCTs: Lead and Second Reviewer

	Reviewer	Preliminaries /5	Introduction /5	Design /5	Sampling /5	Data Collection /5	Ethical /5	Results /5	Discussion /5	Total /40	Total (%)	Difference (pts)	Difference (%)
Ford et al., 2023	Lead	5	5	4	4	4	4	5	5	**36**	**90%**	+2	+5%
	Second	5	5	5	4	5	4	5	5	**38**	**95%**		
Fisher et al., 2021	Lead	4	5	3	3	4	5	4	5	**33**	**82.5%**	+4	+10%
	Second	5	5	4	4	5	4	5	5	**37**	**92.5%**		
Raina et al., 2022	Lead	5	5	4	5	5	4	5	5	**38**	**95%**	+2	+5%
	Second	5	5	5	5	5	5	5	5	**40**	**100%**		
Rioux et al., 2024	Lead	5	5	4	5	5	4	5	5	**38**	**95%**	+1	+2.5%
	Second	5	5	4	5	5	5	5	5	**39**	**97.5%**		
Silverberg et al., 2022	Lead	5	5	5	5	5	4	5	5	**39**	**97.5%**	+1	+2.5%
	Second	5	5	5	5	5	5	5	5	**40**	**100%**		
Ymer et al., 2025	Lead	5	5	4	5	5	5	5	5	**39**	**97.5%**	-1	-2.5%
	Second	5	5	5	4	5	4	5	5	**38**	**95%**		

Note. Maximum possible score per domain = 5; maximum total score = 40. Difference (pts) and Difference (%) = second reviewer score minus lead reviewer score (positive values indicate the second reviewer scored higher). Rows shaded blue indicate a discrepancy of 3 or more points.

Appendix D: STROBE checklist

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	39
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	41
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	42-45
Objectives	3	State specific objectives, including any prespecified hypotheses	45-46
Methods			
Study design	4	Present key elements of study design early in the paper	47
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	47
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	47
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	50
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	50
Bias	9	Describe any efforts to address potential sources of bias	47
Study size	10	Explain how the study size was arrived at	48
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	49
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	52-53
		(*b*) Describe any methods used to examine subgroups and interactions	52
		(*c*) Explain how missing data were addressed	52
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	48
		(*e*) Describe any sensitivity analyses	53
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility,	48

		confirmed eligible, included in the study, completing follow-up, and analysed	
		(b) Give reasons for non-participation at each stage	n/a
		(c) Consider use of a flow diagram	n/a
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	48
		(b) Indicate number of participants with missing data for each variable of interest	n/a
Outcome data	15*	Report numbers of outcome events or summary measures	55-60
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	57-58
		(*b*) Report category boundaries when continuous variables were categorized	55
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	60
Discussion			
Key results	18	Summarise key results with reference to study objectives	65
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	69
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	66-69
Generalisability	21	Discuss the generalisability (external validity) of the study results	70-71
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Not provided

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Appendix E: Final approved MRP proposal

https://osf.io/7deja/files/rw3qn

Appendix F: Ethics Approval Letter

Content on page 98 removed due to confidentiality issues

University of Glasgow

Dr Jessica Fish

MVLS College Ethics Committee

Project Title: Comparing face-to-face and direct-to-home approaches to neuropsychological assessment

Application No: 200250084

The College Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the study.

We are happy therefore to approve the project, subject to the following conditions:

- Project end date as stipulated in original application.
- The data should be held securely for a period of ten years after the completion of the research project, or for longer if specified by the research funder or sponsor, in accordance with the University's Code of Good Practice in Research.
- The research should be carried out only on the sites, and/or groups or datasets as defined in the application.
- Any proposed changes in the protocol should be submitted for reassessment, except when it is necessary to change the protocol to eliminate hazard to the subjects or where the change involves only the administrative aspects of the project. The Ethics Committee should be informed of any such changes.
- You should submit a short end of study report within 3 months of completion.

Yours sincerely,

[redacted]

Dr Ourania Varsou
Senior Lecturer in Anatomy
MVLS College Ethics Committee Vice-Chair

School of Medicine, Dentistry and Nursing
College of Medical, Veterinary and Life Sciences
University of Glasgow
Anatomy Facility | Thomson Building |
University Avenue | Glasgow G12 8QQ
Tel: +44 (0)141 330 7726
Email: ourania.varsou@glasgow.ac.uk

The University of Glasgow, charity number SC004401

Appendix G: Participant Information sheet

https://osf.io/7deja/files/nzb3v

Appendix H: Consent form

https://osf.io/7deja/files/whu5e

Appendix I: Data Analysis Plan

https://osf.io/7deja/files/puxrm

Appendix J: Data Analysis Process

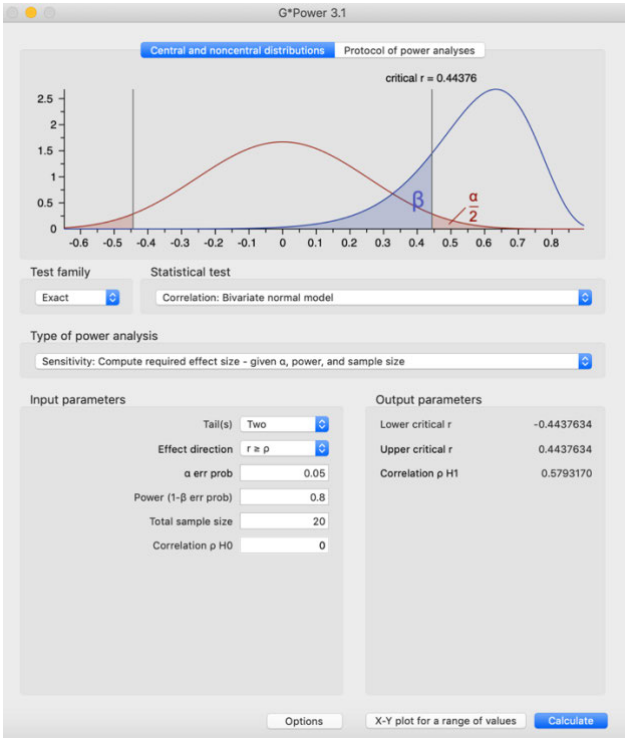
https://osf.io/7deja/files/k2u8y

Appendix K: Data Management Plan (DMP)

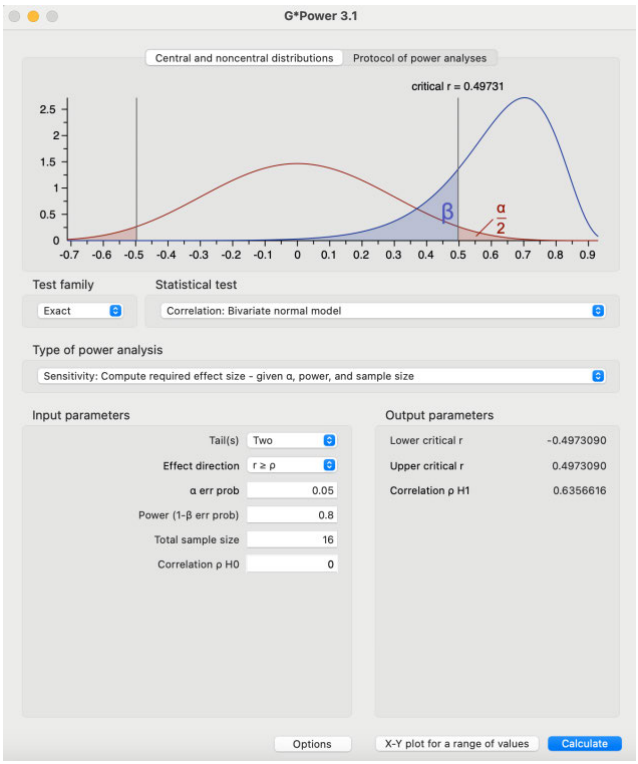
https://osf.io/7deja/files/5qvch

Appendix L: G*Power Sensitivity Analysis for Correlations and Effect Size (Planned Sample, N = 20) and Achieved Sample (N = 16)

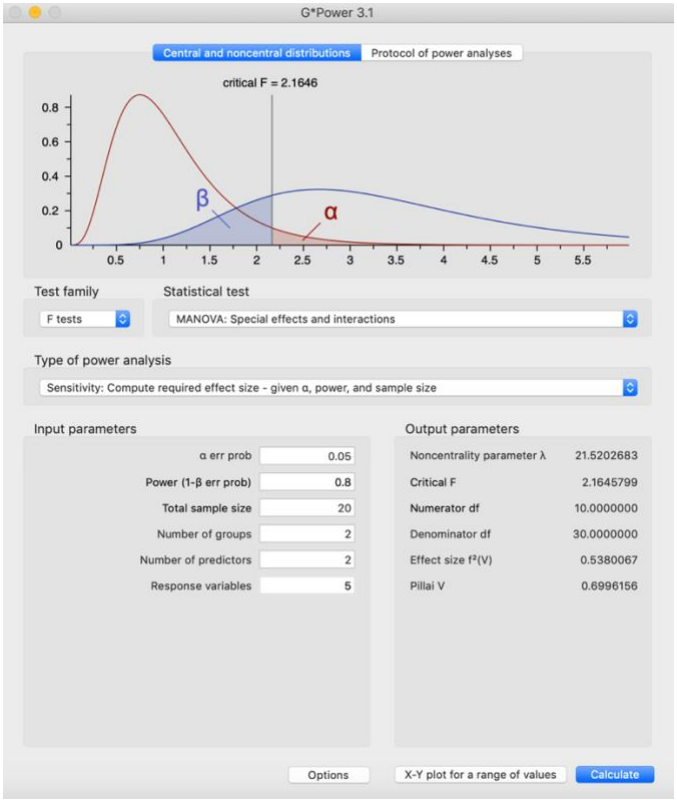
*G*Power Sensitivity Analysis for Correlations (Planned Sample, N = 20)*



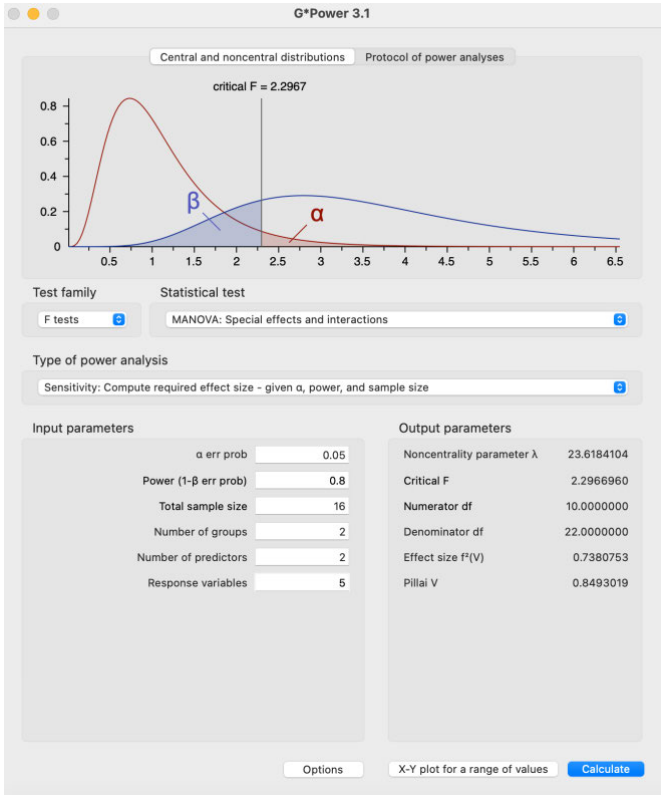
*G*Power Sensitivity Analysis for Correlations (Achieved Sample, N = 16)*



*G*Power Sensitivity Analysis for Effect Size (Planned Sample, N = 20)*



G*Power Sensitivity Analysis for Effect Size (Achieved Sample, N = 16)



Appendix M: Practice Effects Descriptives

Practice Effects Descriptive Statistics for Primary Cognitive Variables:

Session 1 and Session 2 descriptive statistics for all primary cognitive variables

Measure	Session 1 Mdn (IQR) [Min–Max]	Session 2 Mdn (IQR) [Min–Max]	Direction of Change
Verbal Fluency FAS Total	42.00 (13.00) [27–66]	49.50 (21.75) [34–62]	Better
Verbal Fluency Animal Naming Total	21.50 (6.75) [11–32]	23.00 (7.00) [13–31]	Better
RBANS Immediate Memory Index	103.00 (17.25) [73–123]	103.00 (18.00) [81–120]	Same
RBANS Visuospatial Index	101.00 (32.25) [64–131]	109.00 (23.75) [66–121]	Better
RBANS Language Index	98.50 (21.25) [78–114]	100.00 (19.75) [87–120]	Better
RBANS Attention Index	101.50 (20.25) [79–125]	98.50 (14.25) [85–122]	Worse
RBANS Delayed Memory Index	100.00 (17.75) [68–112]	101.00 (18.25) [56–133]	Better
RBANS Total Scale Score Index	99.50 (17.25) [71–113]	101.00 (15.50) [79–121]	Better
Oral TMT-A (z)	−.74 (1.75) [−7.48–.24]	−.93 (1.19) [−5.34–.19]	Worse
Oral TMT-B (z)	−.28 (1.18) [−2.83–.99]	.34 (1.48) [−4.18–1.20]	Better
Digit Span Forward	9.00 (2.00) [4–11]	9.00 (2.00) [6–11]	Same
Digit Span Backward	7.50 (2.00) [4–10]	7.00 (1.75) [4–11]	Worse
DKEFS Colour Naming	9.00 (3.75) [2–12]	9.50 (4.50) [3–13]	Better
DKEFS Word Reading	11.00 (2.50) [5–12]	11.00 (3.50) [5–12]	Same
DKEFS Inhibition	11.00 (2.00) [8–14]	12.00 (4.00) [5–14]	Better
DKEFS Inhibition Switching	10.00 (3.00) [8–12]	11.00 (3.50) [4–14]	Better

Note. Mdn = median; IQR = interquartile range (25^{th} – 75^{th} percentile); Min = minimum; Max = maximum. Direction of change reflects comparison of Session 2 to Session 1 median scores. For Oral TMT-A and Oral TMT-B z scores, more negative values indicate worse performance relative to normative expectations; 'better' therefore reflects less negative or more positive z-scores in Session 2. A mixed pattern of Session 2 changes were observed, with no consistent practice effect across all primary measures.

Appendix N: Condition x Age Group Interaction: Full Univariate Results

Condition x Age Group Interaction: Full Univariate Results (GLM, Tests of Within-Subjects Contrasts)

Univariate tests of the Condition x Age Group Interaction for all 16 primary cognitive measures

Measure	F(1, 11)	p	η^2p
Verbal Fluency FAS Total Correct	0.05	.822	.005
Verbal Fluency Animal Total Correct	0.76	.401	.065
RBANS Immediate Memory Index	0.42	.530	.037
RBANS Visuospatial Index	1.02	.335	.085
RBANS Language Index	1.02	.335	.085
RBANS Attention Index	0.13	.730	.011
RBANS Delayed Memory Index	0.19	.672	.017
RBANS Total Scale Score	0.02	.895	.002
Oral TMT-A (z)	0.08	.787	.007
Oral TMT-B (z)	0.19	.672	.017
Digit Span Forward	0.60	.454	.052
Digit Span Backward	0.09	.772	.008
DKEFS Colour Naming	0.14	.718	.012
DKEFS Word Reading	0.09	.770	.008
DKEFS Inhibition	0.89	.367	.075
DKEFS Inhibition Switching	0.85	.376	.072

Note. None of the 16 interactions were statistically significant.

Appendix O: Sample MRP Reflexivity Journal Entry

https://osf.io/7deja/files/ecajw