A META-ANALYTIC REVIEW EXAMINING THE VALIDITY OF TESTS OF EXECUTIVE FUNCTIONING TO PREDICT FUNCTIONAL, DRIVING AND EMPLOYMENT OUTCOMES IN INDIVIDUALS WITH A TRAUMATIC BRAIN INJURY

And

AN EVALUATION OF THE CONVERGENT VALIDITY OF A FACE-TO-FACE AND VIRTUAL NEUROPSYCHOLOGICAL ASSESSMENT

By

Dr Carl Ryszard Krynicki

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF CLINICAL PSYCHOLOGY

Centre for Applied Psychology
School of Psychology
University of Birmingham
May 2022

Thesis overview

This thesis was submitted in partial fulfilment of the requirements for the degree of Doctor of Clinical Psychology at the University of Birmingham. This thesis presents two studies, the first study is a meta-analytic review examining the validity of tests of executive functioning to predict functional, driving and employment outcomes in individuals with traumatic brain injury. The second paper evaluates the convergent validity of a face-to-face and virtual neuropsychological assessment. The thesis ends with two public dissemination documents for the two research chapters of the thesis.

Dedication

This thesis is dedicated to my family.

Acknowledgements

Firstly, the author would like to thank all of the participants who contributed to the empirical paper and gave up their free time to take part in the study.

The author would also like to acknowledge the outstanding supervision, guidance, patience, and encouragement by Dr Christopher Jones and Dr David Hacker. Both Dr Jones and Dr Hacker have provided exceptional guidance and support throughout his research project, as well as nurturing his interest in working in Clinical Neuropsychology.

Thirdly, the author would like to thank his appraisal tutor, Dr Darrelle Villa, for her ongoing support throughout clinical training and acknowledge the support from the entire course team.

The author would like to thank the outstanding supervision received throughout clinical training. Supervisors include Dr Erik Evenson, Mr Anthony Wynn-Jones, Dr Andrew Sheen, Dr David Hacker, Dr Christopher Jones, Dr Andrew Hawkins, Dr Laura Jordan, Mrs Helen Randall, Mr Jeremy Gunson and the psychoanalytic psychotherapy team, and Dr Michael Ridley-Dash. Additionally, the author would like to acknowledge the hard-working Assistant Psychologists I have had the pleasure of working with whilst on placement, particularly Miss Eyrsa Yasin and Miss Sophie Preece.

Finally, the author would also like to thank the enduring support from his fellow trainees, friends, and family.

Table of Contents

List of Illustrations

List of Tables

CHAPTER I: A META-ANALYTIC REVIEW EXAMINING THE VALIDITY OF TESTS OF EXECUTIVE FUNCTIONING TO PREDICT FUNCTIONAL, DRIVING AND EMPLOYMENT OUTCOMES IN INDIVIDUALS WITH A TRAUMATIC BRAIN INJURY

Abstract

Rationale

Executive functions represent an important domain of abilities that are vital for purposeful goal-directed behaviour. Disruption to these functions is commonly seen in Traumatic Brain Injury (TBI). Previous findings have suggested a link between executive functioning and cognitive recovery, although the evidence for this is mixed. Others have suggested a link between executive functioning and functional outcomes.

Method

A meta-analytic review was conducted. A total of 720 articles were identified from EMBASE, PsychInfo and MEDLINE, and further two articles were hand-searched from references. Twenty-four met inclusion criteria and were included in the review.

Results

The Trail Making Test (part B) (TMT-B) and Wisconsin Card Sorting Test (WCST) were significantly associated with functional outcomes following a TBI. Verbal Fluency was not significantly associated with functional outcome following a TBI. The TMT-B was also associated with a person's ability to return to driving following a TBI, although there were only three studies reporting this outcome. No test of executive functioning was associated with employment outcomes following a TBI, although only four studies were included in this analysis.

Conclusion

Tests of executive functioning, specifically the TMT-B and WCST, were associated with functional outcomes following a TBI, which is important to guide rehabilitation strategies and future planning (such as care needs). This meta-analytic review has also highlighted the scarcity of research in specific outcomes (such as employment and driving).

Introduction

Traumatic brain injury (TBI) is defined as an alteration in brain function or pathology caused by an external force (Menon et al., 2010). TBI remains a leading cause of death and disability, with the most common cause of TBI in young people being road traffic collisions and assault (Lawrence et al., 2016). The mechanical forces of rapid deceleration associated with road traffic collisions and the sharp and abrasive inner surfaces of the skull leaves the frontal and temporal lobes particularly susceptible to trauma, injury to which can result in cognitive and psychosocial difficulties impacting on outcome (Stuss, 2011). TBI is associated with 'executive' dysfunction (Stuss et al., 1985) and there is evidence to suggest that 'executive functions' are associated with the frontal lobes, specifically the prefrontal cortex (Cicerone et al., 2006), but also posterior and subcortical regions (Cristofori et al., 2019). The consequences of TBI can incur not only a burden on the individual and their family but also society (Maas et al., 2017), with an estimated annual cost to society of £15 billion in the UK (Parsonage, 2016).

Executive functions comprise higher top-down regulation of cognitive processes important for engaging in purposeful goal-directed behaviour (Løvstad et al., 2016; Waid-Ebbs et al., 2012), including memory, attention, inhibition, planning and organising, initiation of activity and evaluation (Donders et al., 2015). Five domains of executive functioning have been proposed, and comprise shifting between tasks, updating information in memory, response inhibition, generativity (access of information in long term memory) and fluid reasoning (problem solving) (Fisk & Sharp, 2004; Miyake et al., 2000). Disruption of these functions is commonly seen in neurological and neuropsychological conditions, particularly moderate to severe TBI (Dikmen et al., 1995; Donders et al., 2015; Mazaux et al., 1997; Sigurdardottir et al., 2015) and can be challenging to treat (Cicerone et al., 2006).

Functional recovery refers to an individual's ability to connect with others through social relationships, to engage in activities such as leisure, study or employment and to live independently (Olver et al., 1996). As expected, however, their functional outcome varies as a function of the severity of the TBI sustained (Dikmen et al., 1995). Whilst mild TBI with normal imaging is usually associated with good cognitive recovery at 3 months post-injury (Rohling et al., 2011), moderate to severe TBI has been associated with long-term cognitive impairments (Millis et al., 2001; Schretlen & Shapiro, 2003). However, tests of executive functioning may not directly relate to damage to the frontal lobes. A meta-analytic review of thirty studies (n = 1,269) by Henry and Crawford (2004) found that acquired brain injury (and executive dysfunction as measured by Verbal Fluency performance) was not exclusively associated with frontal lobe injury. Indeed, Axelrod et al. (1996) compared the performance of 356 healthy controls to 343 neurologically impaired patients on the Wisconsin Card Sorting Task, and found that the Wisconsin Card Sorting Task provided an overall modest degree of discrimination between patients and healthy controls, although a modified version of the Wisconsin Card Sorting Task (MWCST) (Nelson, 1976) was associated with functional outcome (Burgess et al., 1998). Therefore these findings indicate that tests of executive function can be used to predict functional outcome rather than specific locus of injury. Moreover, some patients with TBI may perform well on neurocognitive assessments, but perform poorly when subject to more demanding, real-world environments that are dependent on more self-directed behaviour, have more distracting stimuli and may be more emotive (Fisher-Hicks et al., 2021).

Despite the caveat of the frontal lobe paradox, it would be reasonable to assume that cognitive dysfunction, as measured by objective neurocognitive testing, would show a relationship to functional recovery. However, many of the studies focusing on the relationship between cognitive performance and functional outcome following TBI have suggested only a modest

relationship. A meta-analytic review of seven studies by Allanson et al. (2017) found that verbal memory, visuospatial construction and executive functioning was related to functional outcome (measured by the Glasgow Outcome Scale – Extended (GOSE), explaining 31% of the variance. However, this meta-analytic review included only seven studies, and many of the studies included recruited a small sample. Furthermore, whilst studies have found a relationship between cognitive skills including attention, processing speed, memory and executive functions, and functional recovery (Ponsford et al., 2008; Spitz et al., 2012), it remains unclear to what extent executive functioning specifically is able to predict post-injury recovery of global functional abilities and social participation.

There is evidence that cognitive impairment and neurobehavioural symptoms arising from TBI can pose a challenge to an individual's ability to secure meaningful employment. A multicentre prospective cohort study of 134 patients aged fifteen or older with a severe TBI found that only 38% of participants had returned to work (Jourdan et al., 2013). Furthermore Dikmen et al. (1994) found that only 25% of 366 adults returned to work following a severe TBI. Moreover, the most significant barrier to return to work after a severe TBI was reported to be self or other reported cognitive dysfunction compared with behavioural difficulties or physical impairments (Benedictus et al., 2010). There are several studies that have demonstrated the validity of neuropsychological tests to predict employment (Benge et al., 2007; Bercaw et al., 2011; Boake et al., 2001; Cifu et al., 1997).

Potentially, loss of the ability to drive is one of the greatest impediments to return to meaningful activity following traumatic brain injury. The ability to drive relies on the co-ordinated integration of a number of motor, cognitive and perceptual abilities, which include attention, visual information processing, decision making and judgement and shifting between tasks (Schultheis & Whipple, 2014). These may be impaired following a TBI and can significantly

impact on an individual's quality of life, capability to reintegrate into roles and ability to seek employment. Performance on the Trail Making Test (part B) has been found to be associated with returning to driving (Lundqvist et al., 2008; Wolfe & Lehockey, 2016), although the association between Verbal Fluency and driving ability following a TBI has received less attention. Being able to predict whether an individual is likely to be able to drive following a TBI can not only inform treatment and rehabilitation strategies but can also help an individual prepare for re-adjustment if they are not able to return to driving and help the individual plan.

Therefore, the evidence from individual studies indicats that neuropsychological tests have some degree of validity in predicting functional outcome following TBI and can inform rehabilitation strategies and lifestyle adaptations.

To date, however, there has been no specific meta-analytic reviews undertaken of the relationship between executive functioning in TBI and its relationship with functional outcomes. The present review, therefore, aims to undertake a meta-analytic review of the relationship between executive function, as defined by performance on three of the most commonly used measures of executive functioning (Trail Making, Verbal Fluency and the Wisconsin Card Sorting Test), and functional outcomes. Specific outcomes will including global functioning (as measured by the Glasgow Outcome Scale, the Disability Rating Scale, the Mayo Portland Adaptability Index and the Community Integration Questionnaire), employment and driving. Where the data permits sub-analyses will be undertaken to examine the specific relationship of functional outcome to the three individual executive functioning tests.

Primary aims and objectives

There were three overarching aims to this meta-analytic review:

1. To assess the validity of tests of executive functions to predict functional outcome for those with a TBI.
2. To assess the validity of tests of executive functions to predict whether those with a TBI are able to resume driving.
3. To assess the validity of tests of executive functions to predict whether individuals can take up employment/return to employment following a TBI.

Methods

Identifying primary studies

Search of Electronic Databases

A systematic search of the literature was initially carried out in May 2021 using EMBASE, PSYCHINFO, PUBMED and MEDLINE. The aim of the search was to obtain a comprehensive overview of the literature into tests of executive functioning for traumatic brain injury. The search terms that were used to identify these studies are outlined in Table 1 below.

Table 1: Search Criteria

Construct	Free Text Search Terms	Method of Search
Test of executive functioning	"Wisconsin card sorting test" "Trail making test" "Verbal fluency test" "Wisconsin card sorting" ("Trail making test" OR "Verbal fluency test" OR "Wisconsin card sorting") "COWAT" OR "Controlled Oral Word Association Test"	All search terms combined with *OR*
Traumatic brain injury	"Traumatic brain injury" "Brain concussion"	

Outcome	"Recovery (disorders)" "Health outcomes" "Psychosocial outcomes" "prognosis" (recover* OR Outcome* OR prognosis OR "long term" OR "follow up")
Employment	"paid employment" OR "employment" OR "employed" OR "supported employment" OR "return to work"
Driving ability	"drive" OR "driving" OR "driving ability" OR "driving behavior" OR "driving test" OR "automobile driving" OR "automobile driving examination"

Inclusion Criteria

Full inclusion/exclusion criteria are described in Table 2.

Table 2: Inclusion and exclusion criteria.

Inclusion/exclusion criteria	Justification
Nature of intervention: Interventional studies were included in this meta-analysis	Interventional studies that included the relevant outcome data were included in this meta-analysis.
Participant focus Studies that focus on outcome for patients who have experienced a traumatic brain injury	This is to address a gap in the literature of assessing whether assessments of executive functioning can predict outcome for those with a traumatic brain injury.
Outcome data The studies are required to report either Means and Standard Deviations, or F- Test statistics, Cohen's d effect size or an r effect size.	To ensure that outcomes can be calculated into an effect size.
Type of article The following article types were excluded: meta-analysis/theoretical papers/ reviews/commentaries/ clinical guidance/conference abstracts	These articles do not provide the outcome data needed for this meta-analysis.
Outcome Data and study design (N<10, single-case designs, Case series) When the study does not present group data and only provides individual scores.	This is to ensure that an effect size can be calculated and increases methodological rigour of studies included.

The results of the systematic search are presented in *Figure 1*. The search yielded a total of 720 articles and 661 once duplicated were removed. References cited in the papers were hand searched, which yielded two further eligible articles. These articles were then screened using the exclusion criteria using the study titles and abstract. The two most common reasons were: the sample consisted of mainly acquired brain injury (ABI) participants, and tests of executive

functioning were not used to assess outcome. The full text of the remaining seventy-six articles were then reviewed in more detail against the exclusion criteria and twenty-three studies satisfied the criteria for inclusion within this meta-analysis.

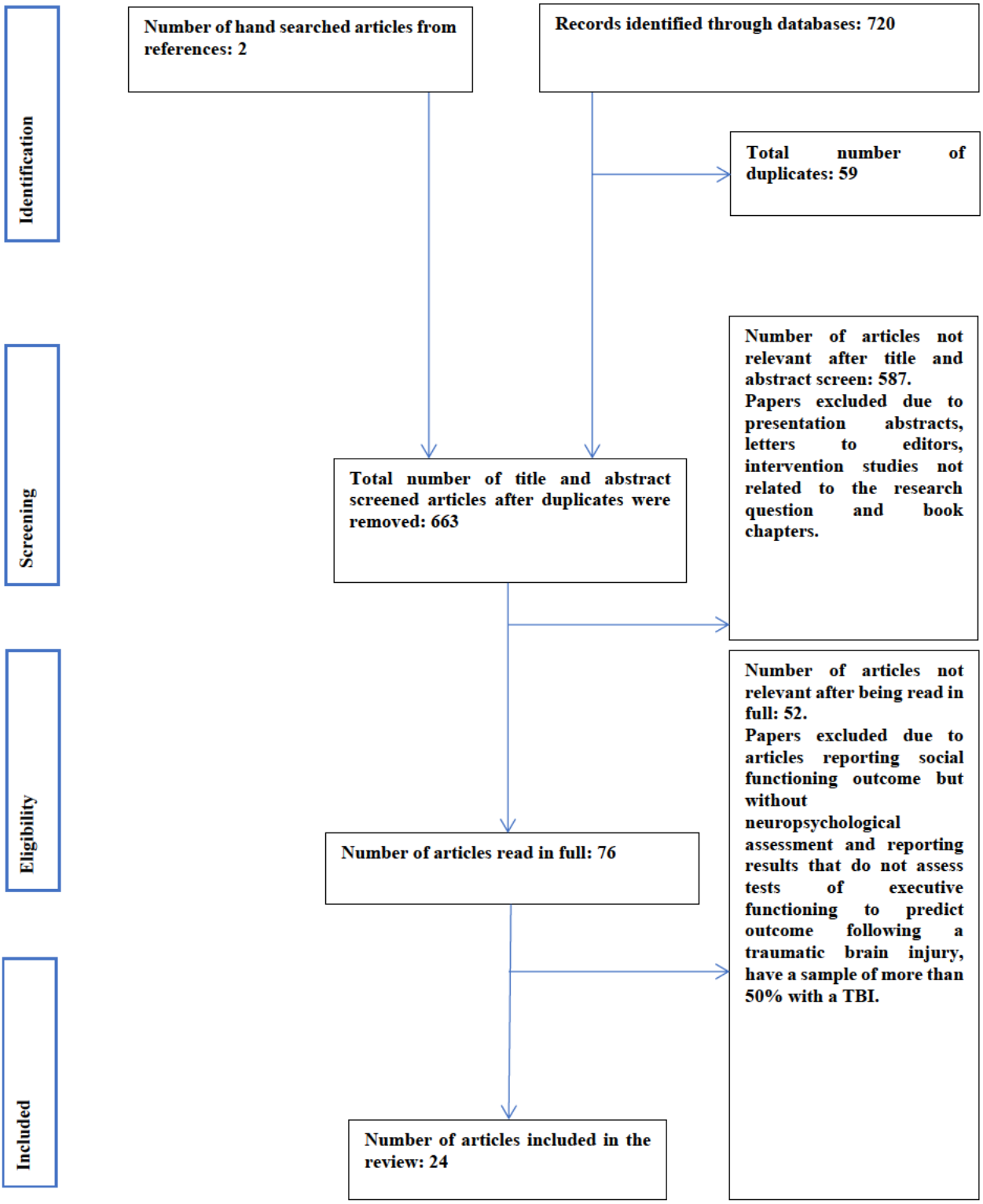


Figure 1: Results of the systematic search and the application of the inclusion criteria

Outcome measures

The assessment of executive functioning is dependent on objective measures of performance on a range of tasks that correspond to executive functions. Commonly used tests to assess executive functioning comprise:

1. The Wisconsin Card Sorting Test (WCST) is a complex problem solving task whereby participants sort cards according to a rule, which then changes without warning. This test assesses cognitive flexibility, memory, attention, inhibition, and planning and organising. Performance on the WCST has been found to be impaired in those with a moderate – severe TBI (Milner, 1971; Ord et al., 2010).
2. The Trail-Making Test (TMT) consists of two parts that must be performed with speed and accuracy; the first part requires participants to connect encircled numbers randomly distributed in sequential order. Part B requires participants to connect numbers and letters in sequential order, whilst alternating between the two, therefore measuring an individual's speed of processing, visual search and shifting between tasks (Arbuthnott & Frank, 2000; Crowe, 1998; Strauss et al., 2006). TMT performance has been found to be sensitive to TBI related impairment (Periáñez et al., 2007).
3. The Verbal Fluency Test (VF) has two parts; firstly, participants are required to produce as many words as possible beginning with a specific letter (phonemic/lexical verbal fluency) and then participants are required to generate words within a specific semantic category, and then switch between two categories. VF is a commonly used test of executive functioning (Amunts et al., 2020) and maps onto the 'generativity' and 'shifting' domains of executive functioning proposed by Miyake et al. (2000).

Moreover, performance has been found to be impaired in those with a TBI (Henry & Crawford, 2004).

In this review, a binary outcome of either employed or unemployed was used to determine the relationship between executive functioning and employment status. Similarly studies reporting whether an individual had returned to driving was used to determine the relationship between executive functioning and return to driving or used the Driving Assessment Scale (Novack et al., 2006) to determine driving ability. Four outcome measures were included in the meta-analysis to assess global functional outcome:

1. The Glasgow Outcome Scale – Extended (GOSE) categorises an individual's level of disability using eight clinician-rated items ranging from good recover to death, and has good test-retest reliability (weighted K coefficient - .92 - .98) (Wilson et al., 2002) and inter-rater reliability (Lu et al., 2010).
2. The Disability Rating Scale (DRS) uses eight clinician-rated items to chart an individual's recovery, and has good internal consistency (α = .83 - .84) and inter-rater reliability (r = .91 - .98) (Malec et al., 2012).
3. The Mayo-Portland Adaptability Inventory (MPAI) (Malec et al., 2007) assesses barriers to community integration through thirty-five questions exploring ability, adjustment, participation, and behaviours following a brain injury and is rated on a 5-point Likert scale, and is shown to have adequate validity and reliability (Kean et al., 2011).
4. The Community Integration Questionnaire (CIQ) (Willer et al., 1993) is a 15-item measure that aims to assess an individual's functioning in three domains; home, social and productivity. The CIQ has been found to provide a valid and reliable measure of and individuals functioning following a TBI (Sander et al., 1999).

Data extraction

It is anticipated that associations will be reported as a zero-order Pearson's *r* correlation coefficient. If treatment outcomes are reported using nonparametric measures of association (e.g., Spearmans Rho, Tau or a Phi coefficient) then the Pearson coefficient will be approximated using the transformations reported by Rupinski and Dunlap (1996). Finally, standardised regression coefficients will be substituted when zero-order correlation coefficients are not reported. Peterson and Brown (2005) have demonstrated that that knowledge of corresponding beta coefficients to impute missing correlations generally produces relatively accurate and precise population effect-size estimates with a meta-analysis. However, it should be noted that regression coefficients as reported in primary studies are frequently calculated from data that has been adjusted for the association with one or more additional covariates. Such adjustments emphasise the idiosyncratic character of the reported regression coefficients and may result in dissimilarity with the effects reported within the other primary studies. The contribution of standardised regression coefficients to overall heterogeneity will be examined empirically if problematic heterogeneity is identified in the random effects model.

Multiple reporting of outcomes can result from primary studies reporting multiple measures of the same outcome or reporting the same outcome measure in multiple subgroups. Where possible multiple outcomes will be combined in a single quantitative outcome using the procedures described by Borenstein et al. (2009). If it is not possible to combine the multiple effects into a single quantitative effect, then the multiple effects will be included in the meta-analysis. The inclusion of multiple reporting of outcomes from that same primary study may result in a slight reduction in confidence intervals for the random effects model as the sample size of that primary study will be included twice.

Defining problematic variance

A study level effect is considered heterogeneous if it presents with variation from the meta-analysis synthesis that cannot be attributed to true variation in the distribution of effect in the population. Heterogeneity can result from methodological variation in the studies, measurement error or uncontrolled individual difference factors within the body of literature. Higgins I^2 is a commonly used measure of heterogeneity, with greater values of I^2 indicating variation in effect that cannot be attributed to true variation in the distribution of effect in the population. As there is considerable variation in methodologies of the primary studies that was used to calculate the meta-analytic synthesis, problematic heterogeneity was defined as a Higgins I^2 value greater than 75%. Where unacceptable or problematic heterogeneity is observed then the focus of the subsequent analyses will be upon the identification of the sources of heterogeneity between the estimates of executive functioning and outcomes in the primary studies.

Risk of Bias Assessment

A set of quality criteria were developed to assess any risk of bias within this literature. The quality criteria were adapted from existing risk of bias frameworks, including The Cochrane Collaboration Risk of Bias Tool (Higgins et al., 2011) and the Risk of Bias Assessment Tool for Nonrandomised Studies (Kim et al., 2013). The current framework assesses risk of bias in seven domains: selection bias, performance bias, treatment fidelity, detection bias, statistical bias, reporting bias and generalisation. The risk of bias in the seven domains and the criteria for Low, Unclear or High risk is described in *Table 3* and the application of these criteria are reported in *Table 5*.

Table 3: Domains of risk of bias and the criteria for ratings of low, unclear, or high risk

Domain	Details	Risk of Bias
Selection Bias	Were efforts made to minimise selection bias in the intervention studies such as consecutive sampling? Was convenience sampling used? If so, studies should potentially be penalised.	High Risk- Includes an unacceptable (reporting less than 30% of the data) level of non-response rate. Target sampling was used. The characteristics of the study population are not reported. Unclear Risk- Non-response rate is not reported. The characteristics of the study population are not clearly reported. For example, the country, setting, location, population demographics were not adequately reported. Further to this, characteristics related to burnout, moral distress and healthcare were not adequately reported e.g. type of occupation, years in service, client group population. The recruitment process/ sampling method of individuals are unclear or has not been reported. Convenience sampling was used. Low Risk- The characteristics of the study population are clearly described and without evidence of bias. Non-response rate is reported and of an acceptable level (set at 50%). The source population is well described, and the study reports the characteristics of the sample e.g. the study details subgroups. The recruitment method is clearly reported and well defined. The article provides some reassurance that there is no selection bias
Performance Bias	Are the outcome measures used valid and reliable for this population?	High Risk- Responses are not confidential or anonymous. Participants were rewarded for their participation in the study. Participants were told which condition/ what questionnaires they were completing and why and any proposed hypotheses. Unclear Risk- The study does not report levels of confidentiality and anonymity. It is not clear if participants were rewarded for their participation (e.g. motivation to respond in a certain way). It is unclear how much information was provided to the participant prior to taking part in the study Low Risk- Study reports level of confidentiality and anonymity. Participants were not rewarded for their participation in the study Information and procedures are provided in a way that does not differentially motivated participants
Treatment Fidelity	Was the treatment sufficiently well described so that it could be replicated?	High Risk – No mention of treatment fidelity tests or processes used to ensure fidelity. Combined with another treatment, no protocol. Unclear Risk – Treatment fidelity undertaken but not described or evaluated. Unclear if following treatment protocol. Training of those delivering the intervention not reported.

Domain	Details	Risk of Bias
		Low Risk – Treatment fidelity described and adequate adherence to model demonstrated. Valid treatment conducted by someone with suitable experience.
Detection Bias	Are the outcome assessors blind to participant allocation?	High Risk - The outcome measures were implemented differently across participants. The outcome measures used had poor reliability and validity reported e.g. Cronbach's Alpha < 0.6. and/or test/retest reliability < 0.6 States that it has been translated but does not detail how this was conducted or clear problems in translation. Only using one dimension/ subscale of the scale or separating the subscales/ dimensions in the analysis. Unclear Risk- Information regarding the outcome measures are either not reported or not clearly reported e.g. definition, validity, reliability. Cronbach's Alpha for outcome measures is between 0.6 and 0.7. Test retest reliability for outcome measures is between .6 and .7 It is not clear if the measure was implemented consistently across all participants. The research question is unclear. Unclear if translated. Low Risk- The outcome measures are clearly defined, valid and reliable, and are implemented consistently across all participants. Outcomes are blindly rated.
Statistical Bias	Have appropriate statistical methods been used? Is there incomplete data due to attrition? Has completer analysis been performed only, or have the studies included an "intention to treat" analysis?	High Risk- Statistics were not reported. Wrong statistical test was used and not appropriate for the study design. Attrition rate – data loss is reported at analysis at an unacceptable level (30%) Unclear Risk – Unclear what statistical test was used. Appropriate statistical test was used but the statistic cannot be transformed into a Pearson's value. Confidence intervals or exact p-values for effect estimates were not reported and could not be calculated. Attrition rate – data loss is not reported at analysis and is therefore unclear Low Risk – Appropriate statistical testing was used. The study has reported a Pearson's value, or the statistic can be transformed into a statistical equivalent. Confidence intervals or exact p-values for effect estimates were given or possible to calculate. Attrition rate – data loss is reported at analysis at an acceptable level (50%)

Domain	Details	Risk of Bias
Reporting Bias	Is there evidence of selective outcome reporting? i.e. only significant results reported. Are there measures that have not been reported in the results that have been mentioned in the method section?	High Risk – Not reported full outcome measures that are stated in the method section/ reported only a subsample of results/only significant results/ not reported the measure as it should be. Unclear Risk – Not all descriptive and/or summary statistics are presented. There is a description (narrative) in the results but do not record statistics. Reported more than one correlation. Low Risk – Reported all results of measures as outlined in the method.
Generalisation	Can the research findings be applied to settings other than that in which they were originally tested? Are there any differences between the study participants and those persons to whom the review is applicable?	High Risk- Small sample with or without idiosyncratic feature (less than 50) High percentage (over 80%) of sample is represented by one professional and cannot be generalised to a variety of healthcare professionals. The sample size is not adequate to detect an effect. Unclear Risk- Sufficient sample for generalisation but with some idiosyncratic features A sample size justification, estimate and power analysis were not provided Low Risk- Sufficient sample for generalisation and representative of target population. (Over 100) A sample size justification, estimate and power analysis was provided. The sample size is adequate to detect an effect

A numeric score was given to a studies' overall risk of bias. A study received two points for a low risk of bias, 1 point for an unclear risk of bias and no points for a high risk of bias in each of the seven risks of bias domains, which was then summed across all of the seven areas of risk of bias. In addition, studies were ranked according to their research design and received additional points according to the quality of the design based on its appropriateness to the research question (see *Table 4*). The overall quality index, reflecting scores for design and the risk of bias, were expressed as a percentage of the total possible score. The overall quality index is reported for each study in the final column of *Table 5*.

Table 4: Study design hierarchy

Study Design	Quality Score	Description
Prospective case cohort study	30	Cohort Study (prospective) is a study of a group of individuals, some of whom are exposed to a variable of interest (e.g., drug or environmental exposure), in which participants are followed up over time to determine who develops the outcome of interest and whether the outcome is associated with the exposure.
Retrospective case cohort study	30	Cohort Study (retrospective) is when data is gathered for a cohort that was formed sometime in the past. Exposures and outcomes have already occurred at the start of the study. You are studying the risk factor and see if you can associate a disease to it. Individuals split by exposure.
Case control study	25	Case Control Study is a study in which patients who already have a specific condition or outcome are compared with people who do not. Researchers look back in time (retrospective) to identify possible exposures. They often rely on medical records and patient recall for data collection.
Cross-sectional studies	20	Cross-Sectional Study is the observation of a defined population at a single point in time or during a specific time interval to examine associations between the outcomes and exposure to interventions. Exposure and outcome are determined simultaneously. Often rely on data originally collected for other purposes.
Before and after study	10	Before and After Study is a study in which within-subject observations are made before (pre) and after (post) the implementation of an intervention/exposure.
Randomised controlled trial/experiment	10	These are experimental studies comparing groups (usually two) to establish the effectiveness of specific interventions The most common design is to compare a new intervention against normal practice (treatment as usual). Participants in the trials are randomly assigned to the treatment groups to minimise bias.
Non-randomised controlled trial/experiment	10	These trials are run when it is not possible to incorporate randomisation into the design. There is an increased risk of biases being introduced into the research and this should be considered carefully when analysis is reported.
Single case experimental study	5	These studies report an interrupted time series within one or a small cohort of participants.
Uncontrolled case study	0	The report of a single case or small cohort without control
Expert opinion	0	Expert opinion

The application of these risk of bias criteria to the characteristics of the primary studies is described in the "traffic light chart" described in the table below.

Study name	Study Design	Selection Bias	Performance Bias	Treatment Fidelity	Detection Bias	Statistical Bias	Reporting Bias	Generalisability	Study Design Score	Risk of bias Score	Overall Quality Score	Overall Quality Index
Cizman-Staba et al (2021)	Cross-sectional studies	Low risk	Unclear risk	Low risk	High risk	Low risk	Low risk	Low risk	20	11	31	70%
Cullen et al (2014)	Prospective case cohort study	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	Low risk	Low risk	30	12	42	95%
de Guise et al (2016)	Cross-sectional studies	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	High risk	20	9	29	66%
Devitt et al (2006)	Prospective case cohort study	Low risk	Unclear risk	Low risk	Low risk	Low risk	Low risk	Low risk	30	13	43	98%
Gautschi et al (2013)	Cross-sectional studies	High risk	Unclear risk	Unclear risk	High risk	Low risk	Low risk	High risk	20	6	26	59%
Goldstein et al (2017)	Randomised controlled trial/experiment	Low risk	Unclear risk	Low risk	Low risk	Low risk	Low risk	Low risk	10	13	23	52%
Hanks et al (2008)	Prospective case cohort study	Low risk	Unclear risk	Low risk	Low risk	Low risk	High risk	Low risk	30	11	41	93%

Hanks, Jackson, et al (2016)	Cross-sectional studies	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	20	12	32	73%
Hart et al (2016)	Randomised controlled trial/experiment	Low risk	Unclear risk	Low risk	High risk	Low risk	Low risk	Low risk	10	11	21	48%
Little et al (1996)	Cross-sectional studies	High risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	20	10	30	68%
Millis et al (1994)	Cross-sectional studies	Low risk	Unclear risk	Low risk	Low risk	Unclear risk	High risk	High risk	20	8	28	64%
Nelson et al (2017)	Prospective case cohort study	High risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	30	10	40	91%
Novack et al (2006)	Cross-sectional studies	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	20	12	32	73%
Ross et al (1996)	Cross-sectional studies	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	20	12	32	73%
Sawamura et al (2018)	Case control study	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	25	12	37	84%
Scott et al (2016)	Prospective case cohort study	Low risk	High risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	30	11	41	93%
Sigurdardottir et al (2009)	Prospective case cohort study	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	High risk	Low risk	30	10	40	91%
Sigurdardottir et al (2018)	Prospective case cohort study	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	30	12	42	95%
Spitz et al (2012)	Prospective case cohort study	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	30	12	42	95%
Vallat-Azouvi et al (2021)	Prospective case cohort study	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	30	12	42	95%
Williams et al (2013)	Prospective case cohort study	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	30	11	41	93%
Wilson et al (2000)	Cross-sectional studies	Low risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	20	12	32	73%
Wilson et al (2021)	Cross-sectional studies	High risk	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	20	10	30	68%
Zafonte et al (2009)	Randomised controlled trial/experiment	Low risk	Unclear risk	Low risk	Low risk	Low risk	Low risk	Low risk	10	13	23	52%

Table 5: Ratings of risk of bias. Red indicates high risk of bias, amber marks an unclear risk of bias and green is a low risk of bias.

Selection Bias

Overall, selection bias was low within the studies. Three studies were rated as unclear risk of bias with four rated as high risk of bias. The low-risk studies either used consecutive sampling and made this clear within their methodology or outline a clear recruitment process (Čizman-Štaba et al., 2021; Devitt et al., 2006; Goldstein et al., 2017; Hanks et al., 2008; Hanks, Rapport, et al., 2016; Hart et al., 2016; Millis et al., 1994; Novack et al., 2006; Ross et al., 1996; Sawamura et al., 2018; Spitz et al., 2012; Vallat-Azouvi et al., 2021; Wilson et al., 2000; Zafonte et al., 2009). The unclear studies sampling methods were often vague, and the recruitment process were often not adequately described.

Performance Bias

Performance bias was unclear with most studies included in this analysis. There was no information regarding any reward for taking part in the study therefore it was unclear if performance in this study was impacted on by an incentive to take part. One study was rated as high performance bias (Scott et al., 2016) due to an incentive being offered for taking part in the study.

Treatment Fidelity

This area of bias was overwhelmingly low. Most studies had other methods in place to ensure validity, such as regular supervision and adequate training of facilitators. All studies outlined the treatment protocol were followed thoroughly or referred to other papers they had written with the protocols described. One study (Gautschi et al., 2013) was scored as unclear risk due to the lack of information regarding supervision and training of research staff.

Detection Bias

The majority of studies did not report on whether those administering outcome assessments were blinded or not and did not report the test-retest statistic. Three studies were rated as high risk of bias in this area because they only used one subscale of the scale in the analysis (Hart et al., 2016) or because it was unclear how they translated the material from English (Čizman-Štaba et al., 2021; Gautschi et al., 2013). The following studies were rated as low risk of detection bias because the outcome data were blindly rated or because the Cronbach Alpha level was above .7 (Cullen et al., 2014; Devitt et al., 2006; Goldstein et al., 2017; Hanks et al., 2008; Zafonte et al., 2009).

Statistical Bias

Twenty-two papers were rated as low risk for this area of bias, with one as unclear and one as high risk. The unclear-risk study had a greater than 20% drop out rate with completer analysis only (Millis et al., 1994).

Reporting Bias

Overall, the full reporting of the outcomes within the studies was considered to be good, with twenty being classed as low risk of reporting bias. Three studies was rated as high risk as some of the test statistics were missing, and therefore could not be used in the analysis (Sigurdardottir et al., 2009) or because some of the analyses were not reported (Hanks et al., 2008; Millis et al., 1994).

Generalisability

Small sample sizes contributed to generalisability being found to be the largest risk area amongst the studies with three being rated as high risk due to having less than fifty participants in the sample (de Guise et al., 2016; Gautschi et al., 2013; Millis et al., 1994). However, given that the majority of studies included in this meta-analysis recruited over 100 participants, the risk of bias for generalisability is low.

Summary

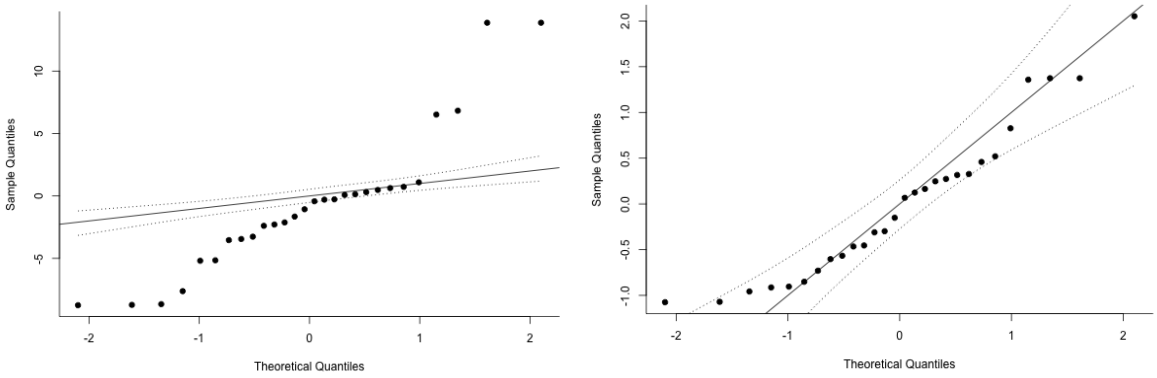
Overall, there was a mixed level of bias across the studies included in the meta-analysis. None of the included studies were without bias. There was a notable unclear risk of bias across studies in the area of performance bias and detection bias domain. Due to the low number of studies in this field, studies with medium to high risk of bias were included. Consequently, the results of this meta-analysis should be interpreted with caution. However, the studies included are felt to be a representative summary of the research literature as it stands currently, and it is hoped that future research will include higher quality research with larger sample sizes.

Results

Functional Outcomes

Selection of the meta-analytic model

The distribution of included study effects is shown in *Figure 2* The between studies variance (tau^2) was calculated using the DerSimonian and Laird estimator (DerSimonian & Laird, 1986).



A *B*

Figure 2: QQ plot of the distribution of DerSimonian and Laird method within the included studies. Plot A shows the distribution of effects relative to the fixed effects model and Plot B depicts the distribution of effects relative to the random effects model using the using the DerSimonian and Laird tau estimate.

Figure 2 depicts the distribution of the study-level effects using the DerSimonian and Laird (DerSimonian & Laird, 1986) method of calculating between study variation (tau). The DerSimonian and Laird (DerSimonian & Laird, 1986) method is the most frequently used method for calculating the between studies variation (tau) when using the random effects model, however this estimator assumes that the effects sizes reported across each of the studies should approximate a normal distribution. As can be seen from *Figure 2*, there is clear evidence of non-normality in the distribution of the correlations when using the fixed effects model, however, this non-normality is largely absent when the random effects model is used. Therefore, the random effects model using the using the DerSimonian and Laird (DerSimonian & Laird, 1986) estimate of between studies variation is appropriate for use with these data.

The omnibus test

The relationship between executive disability and functional outcome is reported in *Figure 3*. There were seventeen studies reporting twenty-eight effects in a total of 5540 participants. Most participants included in this analysis had a moderate – severe TBI (over 21 studies). Two studies included participants with mild, moderate, and severe TBI, and these have been categorised as 'All' in the TBI severity group (5 studies).

Table 6: Study characteristics for studies reporting on functional outcomes

Study	Correlation		95%-CI	% Weight (Random Effects)	Functional outcome measure	Test of executive function	TBI severity group
de Guise et al. 2016.2	0.287	0.0218	0.5522	3.1	MPAI	TMT-B	Mild
Devitt 2006.28	0.0358	-0.0763	0.1479	3.7	CIQ	TMT-B	Moderate - severe
Gautschi et al. 2013.3	0.403	0.0928	0.7132	2.8	GOS-HD	TMT-B	Moderate - severe*
Gautschi et al. 2013.4	0.796	0.6603	0.9317	3.7	GOS-RD	TMT-B	Moderate - severe*
Goldstein 2017.29	0.35	0.2763	0.4237	3.8	GOS	COWAT	Moderate - severe
Goldstein 2017.30	0.33	0.2552	0.4048	3.8	GOS	TMT-B	Moderate - severe
Hanks et al. 2008.11	0.1506	0.005	0.2963	3.6	DRS	TMT-B	Moderate - severe*
Hanks et al. 2016.5	0.0414	-0.034	0.1168	3.8	DRS & GOS	TMT-B	Moderate - severe*
Hanks et al. 2016.6	-0.0003	-0.0759	0.0753	3.8	DRS & GOS	FAS word generation	Moderate - severe*
Hanks et al. 2016.7	-0.0014	-0.077	0.0742	3.8	DRS & GOS	WCST PR	Moderate - severe
Hart 2016.31	0.15	0.0422	0.2578	3.7	GOS	COWAT	Moderate - severe*
Little et al. 1996.12	0.3	0.1742	0.4258	3.7	DRS	WCST PR	All**
Little et al. 1996.13	0.31	0.185	0.435	3.7	DRS	WCST C	All**
Little et al. 1996.14	0.61	0.5232	0.6968	3.8	DRS	TMT-B	All**
Millis 1994.32	0.52	0.1732	0.8668	2.6	CIQ	TMT-B	Moderate - severe
Nelson 2017.33	0.23	0.1388	0.3212	3.8	GOS	TMT-B	All
Novak et al. 2006.16	0.03	-0.2249	0.2849	3.1	DAS	TMT-B	Moderate - severe*
Ross et al. 1997.17	0.3427	0.1156	0.5698	3.2	CIQ	TMT-B	Moderate - severe*
Scott 2016.34	0.18	-0.0909	0.4509	3	MPAI	TMT-B	Mild
Spitz et al. 2012.21	0.0796	-0.0621	0.2213	3.6	MPAI	TMT-B	Moderate - severe
Vallat-Azouvi 2021.23	0.41	0.186	0.634	3.3	GOS	TMT-B	Moderate - severe*
Williams 2013.24	0.0242	-0.0914	0.1399	3.7	DRS	TMT-B	Moderate - severe*

Study	Correlation		95%-CI	% Weight (Random Effects)	Functional outcome measure	Test of executive function	TBI severity group
Williams 2013.25	0.1238	0.0099	0.2377	3.7	DRS	COWAT	Moderate - severe*
Wilson 2000.26	0.35	0.2014	0.4986	3.6	GOS	TMT-B	Moderate - severe*
Wilson 2000.27	0.19	0.0268	0.3532	3.5	GOS	WCST	Moderate - severe
Wilson 2021.37	0.1177	0.0674	0.1681	3.9	GOS	TMT-B	All
Zafonte 2009.35	0.61	0.5665	0.6535	3.9	GOS	TMT-B	Moderate - severe*
Zafonte 2009.36	0.61	0.5665	0.6535	3.9	GOS	COWAT	Moderate - severe*

** Indicates that TBI severity was calculated using the Mayo classification for head injury severity, which determines head injury severity based on trauma imaging findings, loss on consciousness, Glasgow Coma Scale score and post-traumatic amnesia (Malec et al., 2007).*

*** denotes no information given about TBI severity.*

Note: GOS = Glasgow Outcome Scale; GOS – HD = Glasgow Outcome Scale – Hospital Discharge; GOS – RD = Glasgow Outcome Scale – Rehabilitation Discharge; DAS = Driving Assessment Scale; DRS = Disability Rating Scale; MPAI = Mayo Portland Adaptability Inventory; CIQ = Community Integration Questionnaire; TMT – B = Trail Making Test B; COWAT = Controlled Oral Word Association Test; WCST = Wisconsin Card Sorting Test; WCST C = Wisconsin Card Sorting Test (Categories); WCST PR = Wisconsin Card Sorting Test (Perseverative Response).

A random effects models was calculated using the generic inverse variance method. The random effects model for the different levels of TBI severity are shown in *Figure 3*. Significant weighted average correlations were observed for mild TBI (r = 0.2346; 95% CI 0.0451 to 0.4242) and moderate to severe TBI (r = 0.2602; 95% CI 0.1389 to 0.3814). Similarly, a significant weighted average correlation where TBI patients of differing severity had been included in the same dataset (r = 0.3126; 95% CI 0.1215 to 0.5038) and the difference between the TBI severity groups did not reach statistical significance (X^2 = 0.34, p = 0.84). For both the mild TBI and the moderate severe TBI patients, the relationship between test of executive functioning and functional outcomes would be considered of moderate size.

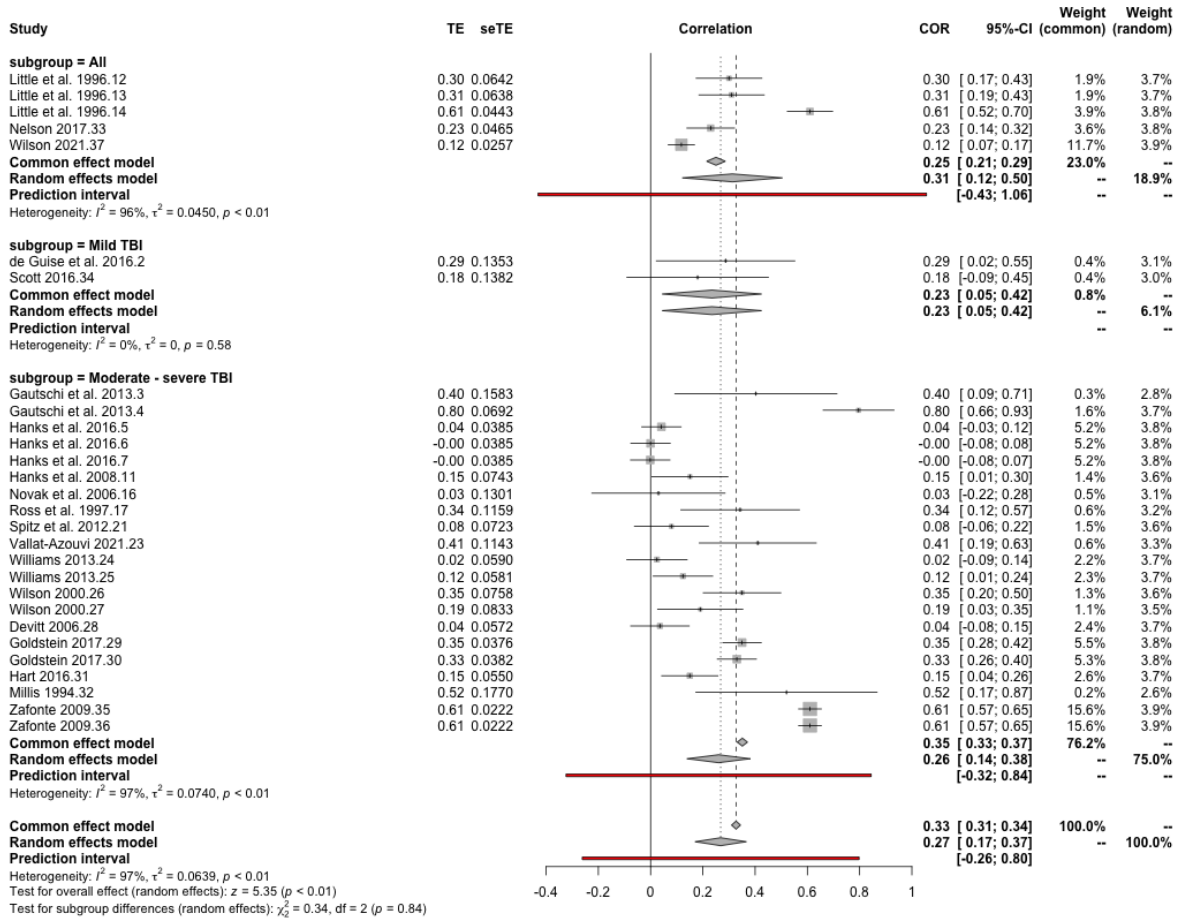


Figure 3: Forest plot of the relationship between executive functioning and functional outcome, separated by TBI severity.

A high level of heterogeneity in the included studies was observed (tau^2 = 0.0639, Higgin's I^2 = 97%; Q = 785.73, p < 0.001), suggesting that the estimates of the association between executive functioning and functional outcome may be biased by the presence of uncontrolled or confounding between studies factors. Therefore, the focus of the subsequent analyses will be upon the identification of the sources of heterogeneity between the estimates of executive functioning and functional outcome.

Impact of executive function test in predicting functional outcome

A random effects models was calculated using the generic inverse variance method. The random effects model for the different tests of executive functioning are shown in Figure 4. Significant weighted average correlations were observed for the Trail Making Test (r = 0.3273;

95% CI 0.3051 to 0.3496) and the Wisconsin Card Sorting Task (r = 0.1344; 95% CI 0.0802 to 0.1887), but not for the Verbal Fluency Task (r = 0.4068; 95% CI 0.3747 to 0.4390). Similarly, a significant weighted average correlation where different tests of executive functioning had been included in the same dataset (r = 0.3308; 95% CI 0.3134 to 0.3481) and the difference between tests of executive functioning did not reach statistical significance (X^2 = 0.80, p = 0.67). The relationship between test of executive functioning and functional outcomes would be considered of moderate size.

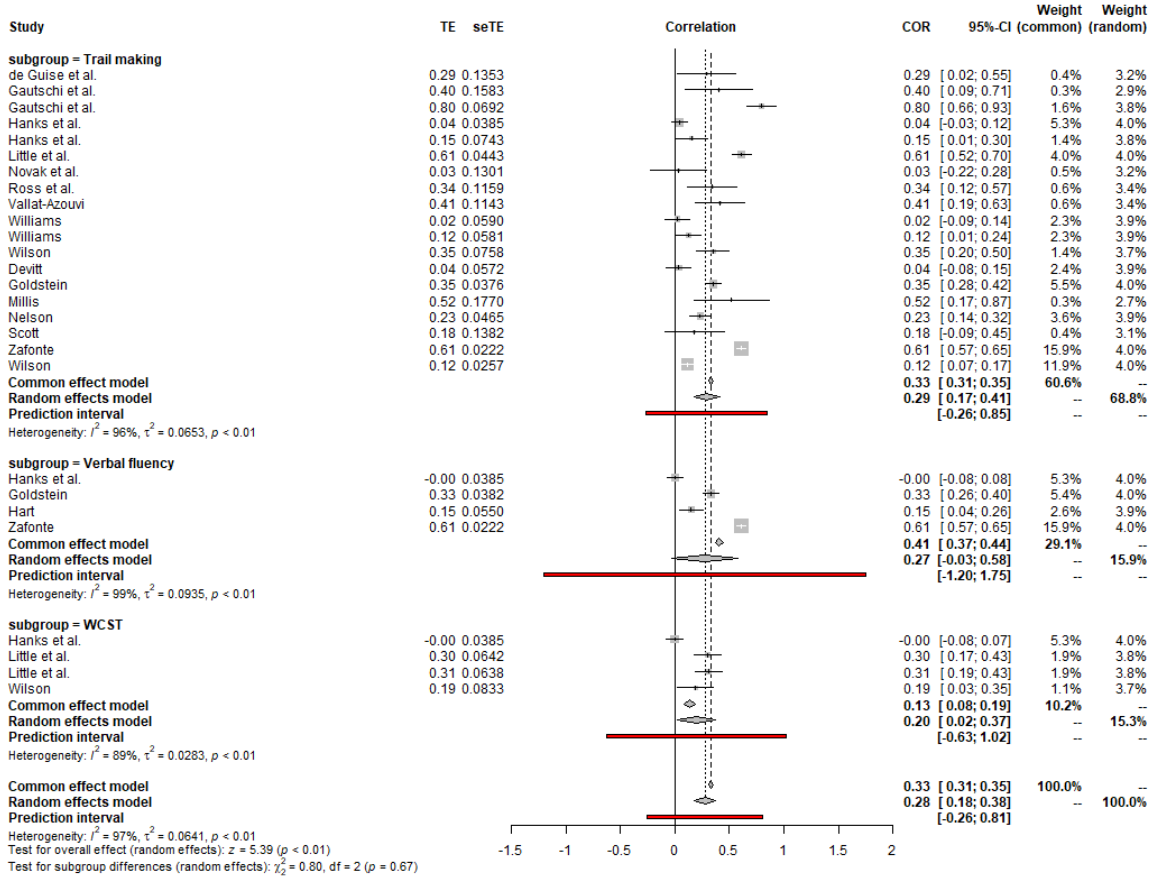


Figure 4: Forest plot showing the relationship between each test of executive functioning and functional outcome

A high level of heterogeneity in the included studies was observed (tau^2 = 0.0641, Higgin's I^2 = 97%; Q = 773.84, p < 0.001), suggesting that the estimates of the association between test of executive functioning and functional outcome may be biased by the presence of uncontrolled or confounding between studies factors.

The impact of influential included studies

The impact of disproportionately influence studies was assessed using a "leave-one-out" analysis, in which the random effects model was calculated with each of the included studies removed in turn and change in weighted average effect size and the change in heterogeneity was recorded. The result of this "leave-one-out" analysis is presented on the Baujat plot (Baujat et al., 2002) in *Figure 5*.

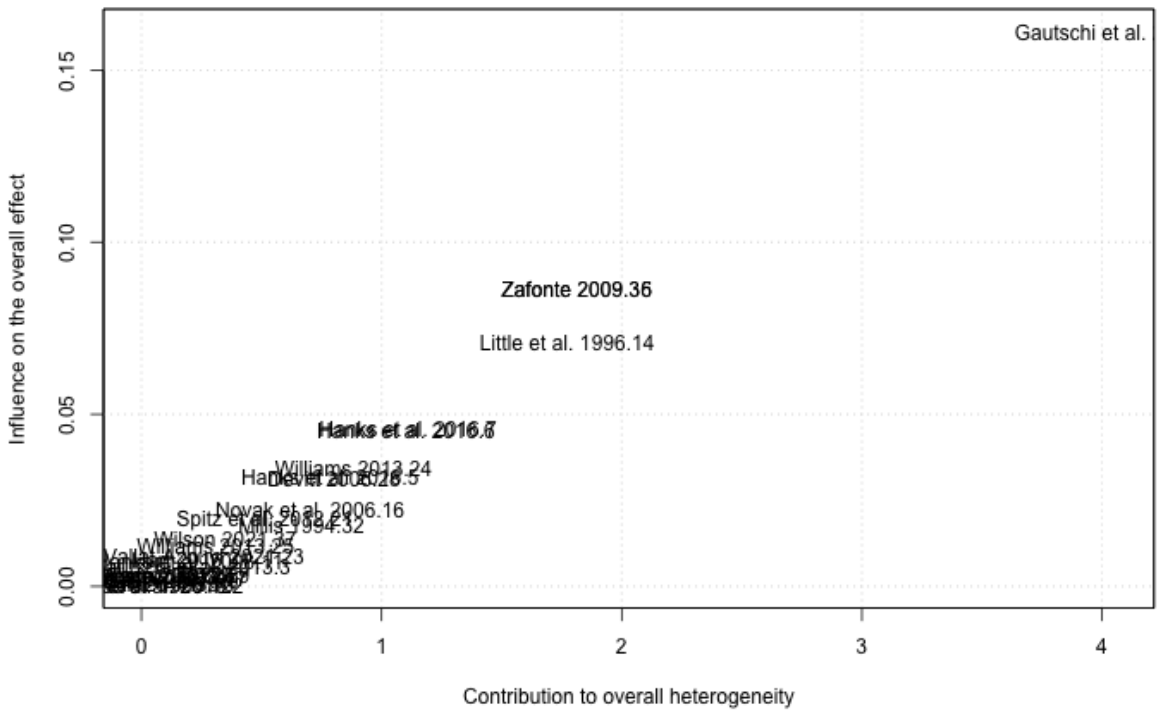


Figure 5: Baujat diagnostic plot of sources of heterogeneity. The vertical axis reports the influence of the study on the overall effect and the horizontal axis reports the discrepancy of the study with the rest of the literature.

As can be seen in *Figure 5*, the effect reported by Gautschi et al. (2013) is both influential and discrepant from the rest of the literature. The random effects model was recalculated with the one study showing disproportionate influence removed. The corrected random effects model reported a synthesis of r = 0.2643 (95% CI 0.1645 to 0.3641). The corrected random effects model evidence only an approximately 0.15% decrease relative to the uncorrected estimate and did not affect the substantive conclusions of this analysis.

The study by Gautschi et al. (2013) was re-reviewed with a view to its removal from this meta-analysis if sufficient concern regarding risk of bias could be identified. As no significant concerns could be identified, therefore, the study by Gautschi et al. (2013) was retained within the meta-analytic synthesis.

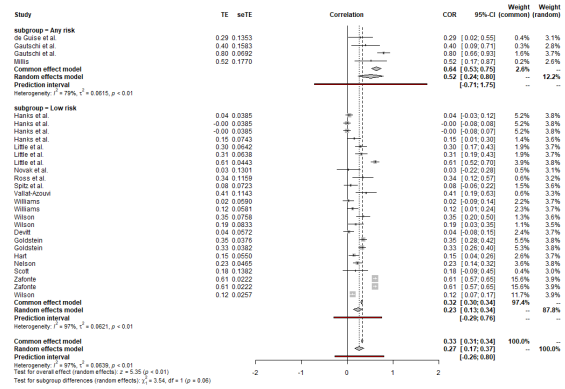
The effect of risk of bias in the included studies

To assess the impact of study level risk of bias upon heterogeneity, a series of subgroup analysis were conducted on the correlation between executive functioning and functional outcomes for the risk of bias ratings of "low risk" and "any risk" (i.e., unclear risk and high risk of bias combined) for each of the seven types of methodological bias.

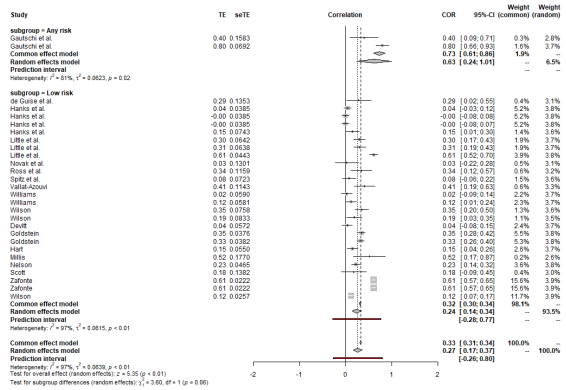
Table 7: Subgroup analysis for the seven areas of risk of bias

	Low Risk			**Any Risk**				
	R	**95% CI**	**k**	**R**	**95% CI**	**k**	**X^2**	**P**
Selection bias	0.2409	0.1118 to 0.3699	18	0.3170	0.1641 to 0.4699	10	0.56	0.4560
Performance bias				0.2682	0.1700 to 0.3664	28	---	---
Treatment bias	0.2435	0.1437 to 0.3432	26	0.6253	0.2435 to 1.0071	2	3.60	0.0579
Detection bias	0.3698	0.2172; 0.5225	7	0.2318	0.1400; 0.3237	21	2.31	0.1289
Statistical bias	0.2657	0.1636; 0.3677	26	0.3004	-0.0550; 0.6558	2	0.03	0.8539
Reporting bias	0.2614	0.1617; 0.3610	27	0.5200	0.1732; 0.8668	1	1.97	0.1601
Generalisability bias	0.2338	0.1304; 0.3371	24	0.5178	0.2406; 0.7951	4	3.54	0.0599

None of the seven areas of risk of bias evidenced statistically significant differences between the studies rated as "low risk" and "any risk", however, treatment bias and generalisability showed trends towards significance (see *Figure 6*).



Generalisability



Treatment Bias

Figure 6: Subgroup plot of generalisability and treatment bias

For both generalisability and treatment bias, low risk of bias was associated with smaller correlations between executive functioning and functional outcomes.

Differences between different functional outcome measures.

A subgroup analysis was undertaken to explore the potential differences between methods and measures of assessing functional outcome. The weighted average effect sizes are shown in *Table 8*.

Table 8: Weighted average correlation between executive functioning and different measures of functional outcome.

Level	R	95% CI	k
Community Integration Questionnaire	0.2676	-0.0251 to 0.5603	3
Disability Rating Scale	0.2551	0.0608 to 0.4493	6
Glasgow-Outcome Scale	0.3785	0.2440 to 0.5130	12
Global Rating Scale	0.03	-0.2249 to 0.2849	1
Mayo-Portland Adaptability Inventory	0.1352	0.0217 to 0.2486	3
	X^2 =	**9.83 p < 0.0434**	

As can be seen from *Table 8*, there was a significant difference between the methods and measures of assessing functional outcome ($X^2 = 9.83$ $p < 0.0434$), with the Global Rating Scale showing lower estimates than the other methods of assessment.

The impact of publication and small study biases

Rosenthal (1979) describes the calculation of a failsafe number; this method calculates the number of with non-significant results which would need to be included in the meta-analysis for the overall effect to be non-significant ($p > .05$). This procedure suggests that 8272 studies would be required to reduce the observed correlation = 0.27 to non-significance, suggesting that the overall conclusions of this meta-analysis are robust to studies missing due to publication bias.

Driving Outcomes

The omnibus test

The relationship between executive disability and driving ability is reported in *Table 9*. There were three studies reporting three effects in a total of 190 participants. Participants were selected from a sample of individuals with a TBI.

Table 9: Study characteristics for studies reporting on driving ability

Study	Correlation		95%-CI	% Weight (Random Effects)	Outcome measure	Test of executive functioning	TBI severity group
Cullen et al.	0.4000	0.1973	0.6027	35.8	Return to driving	TMT B	Moderate - severe
Novak et al.	0.2900	0.0563	0.5237	26.9	Driving assessment scale	TMT B	Moderate - severe
Staba et al.	0.4500	0.2515	0.6485	37.3	Return to driving	TMT B	Moderate - severe

A random effects models was calculated using the generic inverse variance method. The random effects model for the associated between the Trail Making Test and driving ability are shown in *Figure 7*. Significant weighted average correlations were observed (r = 0.3890 95% CI 0.2678 to 0.5103) and the association between the Trail Making Test and driving ability did reach statistical significance (X^2 = 6.29, p < 0.0001).

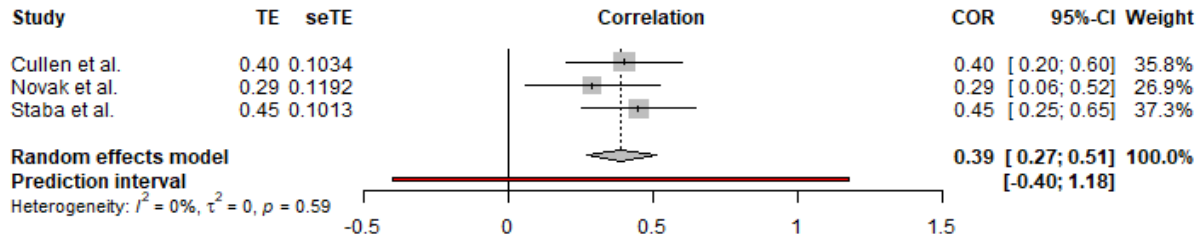


Figure 7: Forest plot of the association between the Trail Making Test and driving ability.

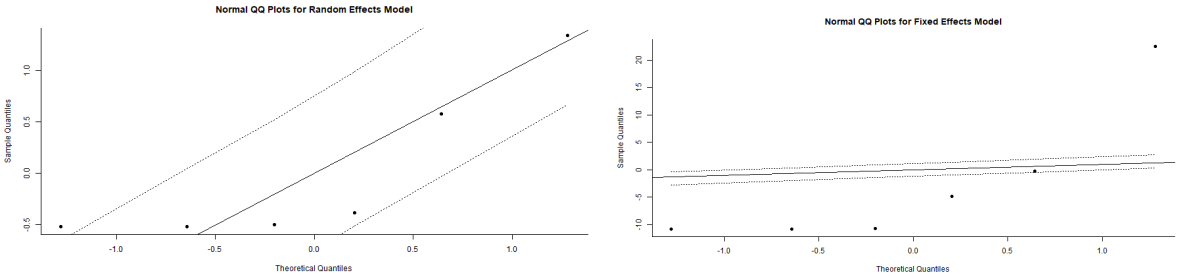
A low level of heterogeneity in the included studies was observed (tau^2 = 0, Higgin's I^2 = 0%; Q = 1.06, p< .05876) (Higgins et al., 2003), suggesting that the estimates of the association between executive functioning and functional outcome may not be biased by the presence of uncontrolled or confounding between studies factors, however, it should be noted that

accurately assessing between studies variation is difficult when there are a small number of primary studies.

Employment Outcomes

Selection of the meta analytic model

The distribution of included study effects is shown in *Figure 8*. The between studies variance (tau^2) was calculated using the DerSimonian and Laird estimator (DerSimonian & Laird, 1986).



A *B*

Figure 8: QQ plot of the distribution of employment outcome within the included studies. Plot A shows the distribution of effects relative to the fixed effects model and Plot B depicts the distribution of effects relative to the random effects model using the using the DerSimonian and Laird tau estimate.

Figure 8 depicts the distribution of the study-level effects using the DerSimonian and Laird (DerSimonian & Laird, 1986) method of calculating between study variation (tau). The DerSimonian and Laird (DerSimonian & Laird, 1986) method is the most frequently used method for calculating the between studies variation (tau) when using the random effects model, however this estimator assumes that the effects sizes reported across each of the studies should approximate a normal distribution. As can be seen from *Figure 8*, there is clear evidence of non-linearity in the distribution of the correlations when using the fixed effects model, however, this non-normality is largely absent when the random effects model is used. Therefore, the random effects model using the using the DerSimonian and Laird (DerSimonian & Laird, 1986) estimate of between studies variation is appropriate for use with these data.

The omnibus test

The relationship between executive disability and employment outcome is reported in *Table 10*. There were four studies reporting six effects in a total of 728 participants. Participants were selected from a homogenous sample of moderate – severe TBI.

Table 10: Study characteristics for studies reporting on employment outcomes.

Study	Correlation		95%-CI	% Weight (Random Effects)	Test of executive function	TBI severity group
Hanks et al. 2016	-0.0001	-0.1012	0.1010	16.7	TMT B	Moderate - severe
Hanks et al. 2016	0.0100	-0.0911	0.1111	16.7	WCST PR	Moderate - severe
Hanks et al. 2016	0.0009	-0.1002	0.1020	16.7	FAS	Moderate - severe
Sawamura et al. 2018	0.8600	0.8154	0.9046	16.9	TMT B	Moderate - severe
Sigurdarottir et al. 2009	0.5114	0.3688	0.6540	16.6	TMT (1 – 5)	Moderate - severe
Sigurdarottir et al. 2018	0.0606	-0.1318	0.2530	16.3	TMT (1 – 5)	Moderate - severe

A random effects models was calculated using the generic inverse variance method. The random effects model for the association between measures of executive functioning and employment status are shown in *Figure 9*. Non-significant weighted average correlations were observed (r = 0.2417 95% CI -0.1676 to 0.6509) and the association between tests of executive functioning and employment status did not reach statistical significance (X^2 = 1.16, p = 0.2471).

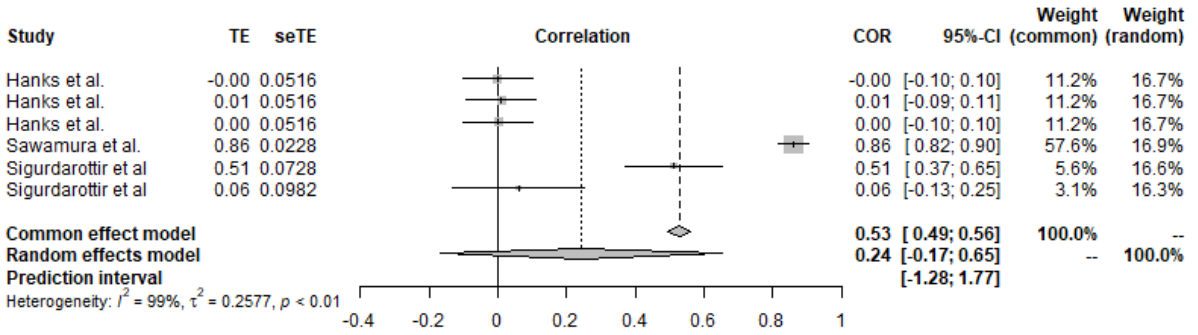


Figure 9: Forest plot of association between the Trail Making Test and employment outcomes.

A high level of heterogeneity in the included studies was observed (tau^2 = 0.2577, Higgin's I^2 = 99%; Q = 545.90, p < 0.0001) (Higgins et al., 2003), suggesting that the estimates of the

association between executive functioning and employment status may be biased by the presence of uncontrolled or confounding between studies factors. It should be noted that it is difficult to accurately assess the source of between studies variation when there are only a few primary studies. However, visual inspection of the forest plot depicted in *Figure 9* would suggest that the studies by Sawamura et al. (2018) and Sigurdardottir et al. (2009) are inconsistent with the small to trivial correlations reported by the other studies. Nevertheless, further studies of this outcome are required before the association between employment status and executive dysfunction can be reliably and accurately assessed.

Discussion

The aims of this meta-analytic review were to assess whether tests of executive functioning were able to predict functional, driving and employment outcomes following a TBI. A further aim was to explore whether this differed by severity of TBI.

This meta-analytic review identified a significant, moderate relationship between test of executive functioning (Trail Making Test and Wisconsin Card Sorting Test) and functional outcome in those with a TBI. However, the high level of heterogeneity required further exploratory analysis to identify the potential sources of heterogeneity, that was not disproportionately influence but outlier studies (such as Gautschi et al. (2013), methodological bias, differences between functional outcome measures (such as the Global Rating Scale (Novack et al., 2006) or publication bias. Further exploration of the data revealed that the association between test of executive functioning and functional outcome was predominantly driven by the Trail Making Test, with the Wisconsin Card Sorting Test and the Verbal Fluency Test corroborating this association. Significantly, the association between test of executive functioning and functional outcome was also confirmed in the mild TBI group, the moderate – severe TBI group and the undifferentiated group of TBI patients. However it is important to

note that the data is correlational, and therefore limiting any direct comparison in TBI severity (i.e. mild versus moderate – severe TBI) and performance on executive functioning test in predicting functional outcome.

A statistically significant, small relationship was identified between the Trail Making Test and driving outcomes in those with a TBI, indicating that the Trail Making Test can reliably predict whether an individual can return to driving following a TBI. Previous research has identified an association between driving and performance on the Trail Making Test (Lundqvist et al., 2008; Wolfe & Lehockey, 2016). However there was a scarcity of research exploring the association between returning to drive following a TBI, with only three studies exploring this association, and only the Trail Making Test (and no other test of executive functioning, such as Verbal Fluency and the Wisconsin Card Sorting Test) being reported as the test of executive functioning in this analysis. The limited number of studies and variety of test of executive functioning included in this analysis as well as the statistically small effect size limits any conclusions drawn from this analysis. Moreover, the three studies that met the studies eligibility criteria for inclusion only reported on a sample of moderate – severe TBI participants, thus limiting the conclusions regarding how transferable these findings may be to a mild TBI population.

When exploring the relationship between tests of executive functioning and employment outcomes, non-significant weighted average correlations were observed indicating that tests of executive functioning cannot reliably predict employment outcomes in those following a TBI. However, only four studies were included in this analysis and heterogeneity was high suggesting that the results may be biased by the presence of uncontrolled or confounding between studies factors. Moreover, the lack of association between employment and executive functioning may be due to the rating of employment as the outcome measure. Hanks, Jackson,

et al. (2016) recorded an individual as employed if they were working either full or part-time, or if they reported they were a home-maker (if this was their position prior to their injury), whereas Sawamura et al. (2018) distinguished between individuals who had achieved competitive employment from those who were in supported employment and those unemployed. These differences in recording employment status may have contributed to the lack of significant associations between tests of executive function and employment outcomes.

This is the first meta-analytic review to examine the specific association between tests of executive functioning and global functional outcome as well as specific functional outcomes of driving and employment. A meta-analytic review by Allanson et al. (2017) explored the neuropsychological predictors of TBI, however this meta-analytic review did not include employment and driving as distinct functional outcomes, only focused on two measures of functional outcome (the Glasgow Outcome Scale – Extended and the Disability Rating Scale), and did not include TBI severity as a variable in the analysis. The results in the current meta-analytic review confirms the findings of Allanson et al. (2017), but considers an individual's TBI severity in the analysis and separated out driving and employment outcomes as separate analyses.

Clinical implications

This review has confirmed that measures of executive functioning are associated with functional outcome in those with either a mild or moderate – severe TBI. Predicting functional recovery following a TBI remains a challenge to clinicians, with some individuals remaining functionally impaired following a TBI despite showing little impairment on formal cognitive assessments (Wilson et al., 2021). Therefore being able to identify which cognitive assessments can adequately predict functional impairment will not only guide rehabilitation strategies but will also help provide some certainty to patients regarding the form their functional recovery

will likely take and whether plans need to be made for future support (such as vocational or domestic).

Previous findings have suggested that executive functioning is an important predictor of functional outcome (Boake et al., 2001; Ponsford et al., 2008), however several other areas of neuropsychological functioning is also important in predicting functional outcome including memory, processing speed and attention (Bercaw et al., 2011; Ponsford et al., 2008), indicating that a full neuropsychological battery of tests is important to fully understand an individual's strengths and weaknesses. The findings from the current review suggests that performance on tests of executive functioning may be used to predict functional impairment.

Being able to predict whether an individual is able to return to driving from tests of executive functioning has important implications for helping to guide rehabilitation and potentially allow for the individual to adjust to a change in their employment circumstances following a TBI. Being able to drive following a TBI, especially in those who were previously able to drive, may significantly impact not only on the individual's quality of life, but can also impact on their employment, home life and psychological wellbeing.

Limitations

One limitation with this meta-analytic review was the limited number of studies focusing on the association between executive functioning and employment, and executive functioning and return to driving, which has prevented any further exploratory analyses to be carried out (such as identifying sources of bias in the data). Moreover, the scarcity of research exploring the association between executive dysfunction and employment outcomes has meant that any conclusions drawn from the lack of association between tests of executive functioning and returning to employment following a TBI, must be drawn tentatively.

A further limitation with this research was the differences in time between the neuropsychological assessment being administered from the TBI being sustained. Some studies reported on participants being assessed six months following a TBI, whereas others assessed participants nineteen years after their TBI. Moreover, some participants had received comprehensive rehabilitation whereas others had not, which may be a confounding variable when considering how well tests of executive functioning can predict functional, driving and employment outcomes following a TBI.

A final limitation with this research is that the majority of tests of executive functioning was the Trail Making Test (part B), which has also been described as a test of information processing speed (Wood & Rutterford, 2006). Therefore, this limits the conclusions that can be made regarding whether tests of executive functioning can be used to predict outcomes or whether the Trail Making Test (part B) has specific explanatory power in predicting outcome following a TBI.

Future directions

Future research should focus on determining whether there is an association between executive functioning and employment outcomes, given that there appears to be a scarcity of research in this important area of clinical outcome. The importance of being able to predict employment outcomes following a TBI has important implications not only for developing rehabilitation strategies, but also for planning support packages which may be required.

The majority of research included in this meta-analytic review utilised the Trail Making Test (part B) to determine the association between executive function and outcome following TBI. One disadvantage with this scale is that it relies on language ability, therefore future research should focus on whether the Colour Trail Making Test can also be used to predict outcome following TBI.

Conclusion

To conclude, this study has highlighted the importance of the association between tests of executive function and functional outcome and tests of executive function and driving outcome. Therefore suggesting that early intervention can be targeted at this important area following a TBI to ensure maximum recovery potential. However, this research has highlighted a dearth of research focusing on executive function and employment outcomes, which is an important focus for future research.

References for meta-analytic review

Allanson, F., Pestell, C., Gignac, G. E., Yeo, Y. X., & Weinborn, M. (2017). Neuropsychological predictors of outcome following traumatic brain injury in adults: a meta-analysis. *Neuropsychology Review, 27*(3), 187-201.

Amunts, J., Camilleri, J. A., Eickhoff, S. B., Heim, S., & Weis, S. (2020, 2020/07/07). Executive functions predict verbal fluency scores in healthy participants. *Scientific Reports, 10*(1), 11141. https://doi.org/10.1038/s41598-020-65525-9

Arbuthnott, K., & Frank, J. (2000). Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. *Journal of Clinical and Experimental Neuropsychology, 22*(4), 518-528.

Axelrod, B. N., Goldman, R. S., Heaton, R. K., Curtiss, G., Thompson, L. L., Chelune, G. J., & Kay, G. G. (1996). Discriminability of the Wisconsin Card Sorting Test using the standardization sample. *Journal of Clinical and Experimental Neuropsychology, 18*(3), 338-342.

Baujat, B., Mahé, C., Pignon, J. P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Statistics in medicine, 21*(18), 2641-2652.

Benedictus, M. R., Spikman, J. M., & van der Naalt, J. (2010). Cognitive and behavioral impairment in traumatic brain injury related to outcome and return to work. *Archives of Physical Medicine and Rehabilitation, 91*(9), 1436-1441.

Benge, J. F., Caroselli, J. S., & Temple, R. O. (2007, Apr 2007). Wisconsin Card Sorting Test: Factor structure and relationship to productivity and supervision needs following severe traumatic brain injury. *Brain injury, 21*(4), 395-400.

Bercaw, E. L., Hanks, R. A., Millis, S. R., & Gola, T. J. (2011, Jan 2011). Changes in neuropsychological performance after traumatic brain injury from inpatient rehabilitation to 1-year follow-up in predicting 2-year functional outcomes. *The Clinical Neuropsychologist, 25*(1), 72-89.

Boake, C., Millis, S. R., High Jr, W. M., Delmonico, R. L., Kreutzer, J. S., Rosenthal, M., Sherer, M., & Ivanhoe, C. B. (2001). Using early neuropsychologic testing to predict long-term productivity outcome from traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 82*(6), 761-768.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4*(6), 547-558.

Cicerone, K., Levin, H., Malec, J., Stuss, D., & Whyte, J. (2006). Cognitive rehabilitation interventions for executive function: moving from bench to bedside in patients with traumatic brain injury. . *J Cogn Neurosci . 18*(7), 1212 - 1222.

Cifu, D. X., Keyser-Marcus, L., Lopez, E., Wehman, P., Kreutzer, J. S., Englander, J., & High, W. (1997). Acute predictors of successful return to work 1 year after traumatic brain injury: a multicenter analysis. *Archives of Physical Medicine and Rehabilitation, 78*(2), 125-131.

Čizman-Štaba, U., Klun, T., & Resnik-Robida, K. (2021). Predicting factors of driving abilities after acquired brain injury through combined neuropsychological and Mediatester driving assessment. *Psihologija, 54*(2), 137-154.

Cristofori, I., Cohen-Zimerman, S., & Grafman, J. (2019). Executive functions. *Handb Clin Neurol, 163,* 197-219. https://doi.org/10.1016/b978-0-12-804281-6.00011-2

Crowe, S. F. (1998). The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the Trail Making Test. *Journal of clinical psychology, 54*(5), 585-591.

Cullen, N., Krakowski, A., & Taggart, C. (2014, Jan 2014). Early neuropsychological tests as correlates of return to driving after traumatic brain injury. *Brain injury, 28*(1), 38-43.

de Guise, E., Bélanger, S., Tinawi, S., Anderson, K., LeBlanc, J., Lamoureux, J., Audrit, H., & Feyz, M. (2016, May 2016). Usefulness of the Rivermead Postconcussion Symptoms Questionnaire and the Trail-Making Test for outcome prediction in patients with mild traumatic brain injury. *Applied Neuropsychology: Adult, 23*(3), 213-222. https://openurl.ebscohost.com/linksvc/linking.aspx?genre=article&issn=2327-9095&volume=23&issue=3&spage=213&date=2016

DerSimonian, R., & Laird, N. (1986, Sep). Meta-analysis in clinical trials. *Control Clin Trials, 7*(3), 177-188. https://doi.org/10.1016/0197-2456(86)90046-2

Devitt, R., Colantonio, A., Dawson, D., Teare, G., Ratcliff, G., & Chase, S. (2006, May 2006). Prediction of long-term occupational performance outcomes for adults after moderate to severe traumatic brain injury. *Disability and Rehabilitation: An International, Multidisciplinary Journal, 28*(9), 547-559.

Dikmen, S. S., Machamer, J. E., Winn, H. R., & Temkin, N. R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology, 9*(1).

Dikmen, S. S., Temkin, N. R., Machamer, J. E., Holubkov, A. L., Fraser, R. T., & Winn, H. R. (1994). Employment following traumatic head injuries. *Archives of Neurology, 51*(2), 177-186.

Donders, J., Oh, Y. I., & J., G. (2015). Self- and Informant Ratings of Executive Functioning After Mild Traumatic Brain Injury. *J Head Trauma Rehabil . 30*(6), E30-E39.

Fisher-Hicks, S., Wood, R. L., & QC, B. B. (2021). The frontal lobe paradox. In *Neuropsychological Aspects of Brain Injury Litigation* (pp. 140-157). Routledge.

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology, 26*(7), 874-890.

Gautschi, O. P., Huser, M. C., Smoll, N. R., Maedler, S., Bednarz, S., von Hessling, A., Lussmann, R., Hildebrandt, G., & Seule, M. A. (2013). Long-term neurological and neuropsychological outcome in patients with severe traumatic brain injury. *Clinical neurology and neurosurgery, 115*(12), 2482-2488.

Goldstein, F. C., Caveney, A. F., Hertzberg, V. S., Silbergleit, R., Yeatts, S. D., Palesch, Y. Y., Levin, H. S., & Wright, D. W. (2017, 2017). Very early administration of progesterone does not improve neuropsychological outcomes in subjects with moderate to severe traumatic brain injury. *Journal of Neurotrauma, 34*(1), 115-120. https://gateway.proquest.com/openurl?ctx_ver=Z39.88-2004&res_id=xri:pqm&rft_val_fmt=ori/fmt:kev:mtx:journal&genre=article&issn=0897-7151&volume=34&issue=1&spage=115&atitle=Very Early Administration of Progesterone Does Not Improve Neuropsychological Outcomes in Subjects with Moderate to Severe Traumatic Brain Injury

Hanks, R. A., Jackson, A. M., & Crisanti, L. K. (2016, Oct 2016). Predictive validity of a brief outpatient neuropsychological battery in individuals 1–25 years post traumatic brain injury. *The Clinical Neuropsychologist, 30*(7), 1074-1086.

Hanks, R. A., Millis, S. R., Ricker, J. H., Giacino, J. T., Nakese-Richardson, R., Frol, A. B., Novack, T. A., Kalmar, K., Sherer, M., & Gordon, W. A. (2008, May 2008). The predictive validity of a brief inpatient neuropsychologic battery for persons with traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 89*(5), 950-957.

Hanks, R. A., Rapport, L. J., Waldron Perrine, B., & Millis, S. R. (2016, Aug 2016). Correlates of resilience in the first 5 years after traumatic brain injury. *Rehabilitation Psychology, 61*(3), 269-276. https://gateway.proquest.com/openurl?ctx_ver=Z39.88-2004&res_id=xri:pqm&req_dat=xri:pqil:pq_clntid=49010&rft_val_fmt=ori/fmt:kev:mtx:journal&genre=article&issn=0090-5550&volume=61&issue=3&spage=269

Hart, T., Novack, T. A., Temkin, N., Barber, J., Dikmen, S. S., Diaz-Arrastia, R., Ricker, J., Hesdorffer, D. C., Jallo, J., & Hsu, N. H. (2016). Duration of post-traumatic amnesia predicts neuropsychological and global outcome in complicated mild traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 31*(6), E1.

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology, 18*(4), 621.

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj, 343*.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj, 327*(7414), 557-560.

Jourdan, C., Bosserelle, V., Azerad, S., Ghout, I., Bayen, E., Aegerter, P., Weiss, J. J., Mateo, J., Lescot, T., Vigué, B., Tazarourte, K., Pradat-Diehl, P., & Azouvi, P. (2013). Predictive factors for 1-year outcome of a cohort of patients with severe traumatic brain injury (TBI): results from the PariS-TBI study. *Brain Inj, 27*(9), 1000-1007. https://doi.org/10.3109/02699052.2013.794971

Kean, J., Malec, J. F., Altman, I. M., & Swick, S. (2011). Rasch measurement analysis of the Mayo-Portland Adaptability Inventory (MPAI-4) in a community-based rehabilitation sample. *Journal of Neurotrauma, 28*(5), 745-753.

Kim, S. Y., Park, J. E., Lee, Y. J., Seo, H.-J., Sheen, S.-S., Hahn, S., Jang, B.-H., & Son, H.-J. (2013). Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *Journal of clinical epidemiology, 66*(4), 408-414.

Lawrence, T., Helmy, A., Bouamra, O., Woodford, M., Lecky, F., & Hutchinson, P. J. (2016). Traumatic brain injury in England and Wales: prospective audit of epidemiology, complications and standardised mortality. *BMJ open, 6*(11), e012197-e012197. https://doi.org/10.1136/bmjopen-2016-012197

Løvstad, M., Sigurdardottir, S., Andersson, S., Grane , V. A., Moberget , T., Stubberud, J., & Solbakk, A. K. (2016). Behavior Rating Inventory of Executive Function Adult Version in Patients with Neurological and Neuropsychiatric Conditions: Symptom Levels and Relationship to Emotional Distress. *J Int Neuropsychol Soc, 22*(6), 682 - 694.

Lu, J., Marmarou, A., Lapane, K., Turf, E., & Wilson, L. (2010). A method for reducing misclassification in the extended Glasgow Outcome Score. *Journal of Neurotrauma, 27*(5), 843-852.

Lundqvist, A., Alinder, J., & Rönnberg, J. (2008). Factors influencing driving 10 years after brain injury. *Brain injury, 22*(4), 295-304.

Maas, A. I., Menon, D. K., Adelson, P. D., Andelic, N., Bell, M. J., Belli, A., Bragge, P., Brazinova, A., Büki, A., & Chesnut, R. M. (2017). Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology, 16*(12), 987-1048.

Malec, J. F., Brown, A. W., Leibson, C. L., Flaada, J. T., Mandrekar, J. N., Diehl, N. N., & Perkins, P. K. (2007). The mayo classification system for traumatic brain injury severity. *Journal of Neurotrauma, 24*(9), 1417-1424. . https://doi.org/doi: 10.1089/neu.2006.0245.

Malec, J. F., Hammond, F. M., Giacino, J. T., Whyte, J., & Wright, J. (2012). Structured interview to improve the reliability and psychometric integrity of the Disability Rating Scale. *Archives of Physical Medicine and Rehabilitation, 93*(9), 1603-1608.

Mazaux, J. M., Masson, F., Levin, H. S., Alaoui, P., Maurette, P., & Barat, M. (1997). Long-term neuropsychological outcome and loss of social autonomy after traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 78*, 1316 - 1320.

Menon, D. K., Schwab, K., Wright, D. W., & Maas, A. I. (2010). Demographics and clinical assessment working group of the International and interagency initiative toward common data elements for research on traumatic brain injury and psychological health. Position statement: definition of traumatic brain injury. *Arch Phys Med Rehabil, 91*(11), 1637-1640.

Millis, S. R., Rosenthal, M., & Lourie, I. F. (1994, Dec 1994). Predicting community integration after traumatic brain injury with neuropsychological measures. *International Journal of Neuroscience, 79*(3-4), 165-167.

Millis, S. R., Rosenthal, M., Novack, T. A., Sherer, M., Nick, T. G., Kreutzer, J. S., High, W. M., Jr., & Ricker, J. H. (2001). Long-Term Neuropsychological Outcome After Traumatic Brain Injury. *The Journal of Head Trauma Rehabilitation, 16*(4). https://journals.lww.com/headtraumarehab/Fulltext/2001/08000/Long_Term_Neuropsychological_Outcome_After.5.aspx

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British medical bulletin*.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology, 41*(1), 49-100.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12*(4), 313-324.

Novack, T. A., Baños, J. H., Alderson, A. L., Schneider, J. J., Weed, W., Blankenship, J., & Salisbury, D. (2006). UFOV performance and driving ability following traumatic brain injury. *Brain injury, 20*(5), 455-461.

Olver, J. H., Ponsford, J. L., & Curran, C. A. (1996). Outcome following traumatic brain injury: a comparison between 2 and 5 years after injury. *Brain injury, 10*(11), 841-848.

Ord, J. S., Greve, K. W., Bianchini, K. J., & Aguerrevere, L. E. (2010, 2010/02/03). Executive dysfunction in traumatic brain injury: The effects of injury severity and effort on the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology, 32*(2), 132-140. https://doi.org/10.1080/13803390902858874

Parsonage, M. (2016). *Traumatic brain injury and offending: An economic analysis*.

Periáñez, J., Ríos-Lago, M., Rodríguez-Sánchez, J., Adrover-Roig, D., Sánchez-Cubillo, I., Crespo-Facorro, B., Quemada, J., & Barceló, F. (2007). Trail Making Test in traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and normative data. *Archives of Clinical Neuropsychology, 22*(4), 433-447.

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*(1), 175.

Ponsford, J., Draper, K., & Schönberger, M. (2008). Functional outcome 10 years after traumatic brain injury: Its relationship with demographic, injury severity, cognitive and emotional status. *Journal of the International Neuropsychological Society, 14*(2), 233-242.

Rohling, M. L., Binder, L. M., Demakis, G. J., Larrabee, G. J., Ploetz, D. M., & Langhinrichsen-Rohling, J. (2011, 2011/05/01). A Meta-Analysis of Neuropsychological Outcome After Mild Traumatic Brain Injury: Re-analyses and Reconsiderations of Binder et al. , Frencham et al. , and Pertab et al. *The Clinical Neuropsychologist, 25*(4), 608-623. https://doi.org/10.1080/13854046.2011.565076

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychol Bull, 86*(3), 638.

Ross, S. R., Millis, S. R., & Rosenthal, M. (1996). Neuropsychological prediction of psychosocial outcome after traumatic brain injury. *Applied Neuropsychology, 4*(3), 165-170.

Rupinski, M. T., & Dunlap, W. P. (1996). Approximating Pearson product-moment correlations from Kendall's tau and Spearman's rho. *Educational and Psychological Measurement, 56*(3), 419-429.

Sander, A. M., Fuchs, K. L., High, W. M., Jr., Hall, K. M., Kreutzer, J. S., & Rosenthal, M. (1999, Oct). The Community Integration Questionnaire revisited: an assessment of factor structure and validity. *Arch Phys Med Rehabil, 80*(10), 1303-1308. https://doi.org/10.1016/s0003-9993(99)90034-5

Sawamura, D., Ikoma, K., Ogawa, K., & Sakai, S. (2018). Clinical utility of neuropsychological tests for employment outcomes in persons with cognitive impairment after moderate to severe traumatic brain injury. *Brain injury, 32*(13-14), 1670-1677.

Schretlen, D. J., & Shapiro, A. M. (2003, 2003/11/01). A quantitative review of the effects of traumatic brain injury on cognitive functioning. *International Review of Psychiatry, 15*(4), 341-349. https://doi.org/10.1080/09540260310001606728

Schultheis, M. T., & Whipple, E. K. (2014). Driving after traumatic brain injury: evaluation and rehabilitation interventions. *Current physical medicine and rehabilitation reports, 2*(3), 176-183.

Scott, K. L., Strong, C.-A. H., Gorter, B., & Donders, J. (2016, Jan 2016). Predictors of post-concussion rehabilitation outcomes at three-month follow-up. *The Clinical Neuropsychologist, 30*(1), 66-81.

Sigurdardottir, S., Andelic, N., Roe, C., & Schanke, A.-K. (2009, Sep 2009). Cognitive recovery and predictors of functional outcome 1 year after traumatic brain injury. *Journal of the International Neuropsychological Society, 15*(5), 740-750. https://gateway.proquest.com/openurl?ctx_ver=Z39.88-2004&res_id=xri:pqm&rft_val_fmt=ori/fmt:kev:mtx:journal&genre=article&issn=1355-6177&volume=15&issue=5&spage=740&atitle=Cognitive recovery and predictors of functional outcome 1 year after traumatic brain injury

Sigurdardottir, S., Andelic, N., Wehling, E., Roe, C., Anke, A., Skandsen, T., Holthe, O. O., Jerstad, T., Aslaksen, P. M., & Schanke, A.-K. (2015). Neuropsychological functioning in a national cohort of severe traumatic brain injury: Demographic and acute injury-related predictors. *The Journal of Head Trauma Rehabilitation, 30*(2), E1. https://bora.uib.no/bora-xmlui/bitstream/1956/17537/4/Sigurdardottir%2bet%2bal.%2b2015.%2bPostprint.%2bNeuropsychological%2bFunctioning%2bin%2ba%2bnational%2bcohort%2bof%2bsevere%2btraumatic%2bbrain%2binjury.pdf

Spitz, G., Ponsford, J. L., Rudzki, D., & Maller, J. J. (2012, Sep 2012). Association between cognitive performance and functional outcome following traumatic brain injury: A longitudinal multilevel examination. *Neuropsychology, 26*(5), 604-612. https://gateway.proquest.com/openurl?ctx_ver=Z39.88-2004&res_id=xri:pqm&req_dat=xri:pqil:pq_clntid=49010&rft_val_fmt=ori/fmt:kev:mtx:journal&genre=article&issn=0894-4105&volume=26&issue=5&spage=604

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. American chemical society.

Stuss, D., Ely, P., Hugenholtz, H., Richard, M., LaRochelle, S., Poirier, C., & Bell, I. (1985). Subtle neuropsychological deficits in patients with good recovery after closed head injury. *Neurosurgery, 17*(1), 41-47.

Stuss, D. T. (2011). Traumatic brain injury: relation to executive dysfunction and the frontal lobes. *Current Opinion in Neurology, 24*(6), 584-589. https://doi.org/10.1097/WCO.0b013e32834c7eb9

Vallat-Azouvi, C., Swaenepoël, M., Ruet, A., Bayen, E., Ghout, I., Nelson, G., Pradat-Diehl, P., Meaude, L., Aegerter, P., & Charanton, J. (2021). Relationships between neuropsychological impairments and functional outcome eight years after severe traumatic brain injury: Results from the PariS-TBI study. *Brain injury*, *35*(9), 1001-1010.

Waid-Ebbs, J. K., Wen, P.-S., Heaton, S. C., Donovan, N. J., & Velozo, C. (2012, Dec 2012). The item level psychometrics of the behaviour rating inventory of executive function-adult (brief-a) in a TBI sample. *Brain injury*, *26*(13-14), 1646-1657.

Willer, B., Rosenthal, M., Kreutzer, J. S., Gordon, W. A., & Rempel, R. (1993). Assessment of community integration following rehabilitation for traumatic brain injury. *The Journal of Head Trauma Rehabilitation*.

Wilson, J., Edwards, P., Fiddes, H., Stewart, E., & Teasdale, G. M. (2002). Reliability of postal questionnaires for the Glasgow Outcome Scale. *Journal of Neurotrauma*, *19*(9), 999-1005.

Wilson, J., Pettigrew, L., & Teasdale, G. (2000). Emotional and cognitive consequences of head injury in relation to the Glasgow Outcome Scale. *Journal of Neurology, Neurosurgery & Psychiatry*, *69*(2), 204-209.

Wilson, L., Horton, L., Kunzmann, K., Sahakian, B. J., Newcombe, V. F., Stamatakis, E. A., Von Steinbuechel, N., Cunitz, K., Covic, A., & Maas, A. (2021). Understanding the relationship between cognitive performance and function in daily life after traumatic brain injury. *Journal of Neurology, Neurosurgery & Psychiatry*, *92*(4), 407-417.

Wolfe, P. L., & Lehockey, K. A. (2016). Neuropsychological assessment of driving capacity. *Archives of Clinical Neuropsychology*, *31*(6), 517-529.

Wood, R. L. L., & Rutterford, N. A. (2006, May 2006). Demographic and cognitive predictors of long-term psychosocial outcome following traumatic brain injury. *Journal of the International Neuropsychological Society*, *12*(3), 350-358. https://gateway.proquest.com/openurl?ctx_ver=Z39.88-

2004&res_id=xri:pqm&rft_val_fmt=ori/fmt:kev:mtx:journal&genre=article&issn=1355-6177&volume=12&issue=3&spage=350&atitle=Demographic

Zafonte, R., Friedewald, W. T., Lee, S. M., Levin, B., Diaz-Arrastia, R., Ansel, B., Eisenberg, H., Timmons, S. D., Temkin, N., Novack, T., Ricker, J., Merchant, R., & Jallo, J. (2009, Dec 2009). The Citicoline Brain Injury Treatment (COBRIT) trial: Design and methods. *Journal of Neurotrauma*, *26*(12), 2207-2216.

CHAPTER II: AN EVALUATION OF THE CONVERGENT VALIDITY OF FACE TO FACE AND VIRTUAL NEUROPSYCHOLOGICAL ASSESSMENT

Abstract

Rationale

The COVID-19 pandemic has highlighted the need for further research evaluating the validity of conducting a battery of neuropsychological assessments virtually compared with face-to-face administration. Previous research has suggested that some neuropsychological assessments yield valid results when administered virtually, however much of the previous research focused on older adults.

Method

A within-subjects, counter balanced design was employed to assess 28 healthy participants. Participants completed a neuropsychological assessment battery covering tests of general intellectual functioning, memory and attention, executive functioning, language and information processing speed, as well as effort.

Results

There was no significant difference between face-to-face administration of the neuropsychological battery compared with virtual administration for the majority of the tests used. However, there were significant differences in the DKEFS Colour Naming Task, with participants making fewer errors on the colour naming task and inhibition/switching task when administered virtually compared to face-to-face administration. There was also a significant age cohort effect in the inhibition/switching task. There was also a trending significant difference in mode of administration for the DKEFS Verbal Fluency Task.

Conclusion

Virtually administered neuropsychological assessments largely provide a valid alternative to face-to-face assessments, however consideration must be given to test selection as well as the population of participants that are being assessed. Other important considerations must focus on preserving the security and integrity of test materials, as well as administration in a medico-legal setting. Future research should focus on validating assessments with specific patient populations and developing a neuropsychological assessment battery using information technology.

Introduction

The use of videoconferencing software (telemedicine) has seen a rapid growth in recent years, and particularly since the outbreak of the COVID-19 pandemic, announced by the World Health Organisation in March 2020 (World Health Organization, 2020). In response to the dangers posed by transmitting COVID-19, many face-to-face clinic and outpatient departments were suspended, and there was an increase in the use of videoconferencing software to continue patient care. Some departments continued to offer a face-to-face service, however, maintaining a face-to-face service while adapting to many of the restrictions imposed raised significant challenges (such as practical difficulties when conducting a neuropsychological assessment and increased patient anxiety) (Coetzer, 2020). Others however, made use of telecommunication software which enabled the clinician to administer a neuropsychological assessment battery without the need of being in the same room as the patient. The use of telemedicine also has the added benefit of allowing individuals to remain in a setting that is comfortable to them for their neuropsychological assessment, which has been found to be favourable option for assessment, particularly for those in rural locations (Hilty et al., 2007; Norman, 2006), and has been found to benefit those with limited mobility (Grosch et al., 2011).

The British Psychological Society (BPS) Division of Neuropsychology (DON) has supported the use of remote technology when undertaking clinical neuropsychological work and has made recommendations regarding the technical considerations that must be considered when conducting a neuropsychological assessment, such as appropriate internet speeds, access to technology and appropriate selection of neuropsychological assessments that may be easily performed online and yield comparable results if administered face-to-face (Bunnage et al., 2020). These recommendations are broadly in line with the recommendations made by the Inter Organisational Practice committee (Bilder et al., 2020) who has made recommendations

regarding specific neuropsychological assessments that may be used in a virtual neuropsychological assessment. They also acknowledge the challenges that face individuals facing a virtual neuropsychological assessment such as unfamiliarity with the appropriate technology required for the assessment and selecting an appropriate platform to host the virtual assessment.

Neuropsychological assessments rely heavily on the use of a battery of tests to assess domains of functioning, such as executive functioning, memory, orientation, attention, and language. Executive functions are important for engaging in purposeful goal-directed behaviour and comprise higher top-down regulation of cognitive processes (Løvstad et al., 2016; Waid-Ebbs et al., 2012). Executive functions are composed of cognitive domains including attention, inhibition, planning and organising, initiation of activity and evaluation (Donders et al., 2015). Memory consists of several functions that help to support the acquisition, retention and retrieval of information and is usually categorised into short-term/working memory and long-term memory (Arciniegas et al., 2013). Language is a unique human ability, which involves the ability to express language, but also includes reading and writing (Arciniegas et al., 2013).

Disruption of these functions are commonly seen in neurological and neuropsychological conditions such as moderate to severe TBI (Dikmen et al., 1995; Donders et al., 2015; Mazaux et al., 1997; Sigurdardottir et al., 2015), neurodegenerative conditions such as dementia (Bondi et al., 2009), stroke (Sinanović, 2010), and tumours (Wefel et al., 2018), and may be challenging to manage (Cicerone et al., 2006). Therefore, reliably assessing an individual's cognitive functioning and evaluating the extent of an individual's neuropsychological abilities were compromised following neurological injury or illness is essential for planning treatment approaches aiding recovery.

The use of video teleconferencing software has enabled many neuropsychological assessments to be administered without the need for the patient to be in the same room as the assessor, however the extent to which these assessments produce valid results remains unknown. Munro Cullum and Grosch (2013) assessed 83 adults with cognitive impairment and 119 healthy controls in a counterbalanced design to establish whether a video teleconferencing-based neuropsychological assessment yielded valid and reliable results compared to standard face-to-face administration. They focused on the Mini Mental State Examination (MMSE) (Folstein et al., 1975), Hopkins Verbal Learning Test – Revised (Benedict et al., 1998), Digit Span forward and backward (Wechsler, 2008), short form Boston Naming Test (Kaplan et al., 1983), Letter and Category Fluency (Delis et al., 2001), and Clock Drawing (Agrell & Dehlin, 1998). They concluded that administering these assessments using telecommunication software provided a valid alternative to face-to-face assessments.

A study of 150 adults by Gnassounou et al. (2021) also found no difference between face-to-face and virtual administration of a brief neuropsychological assessment battery composed of the MMSE (Folstein et al., 1975), Free and Cued Selective Reminding Test (FCSRT) (French version) (Van der Linden et al., 2004), Mahieux Gestural Praxis Battery (Mahieux-Laurent et al., 2009), Frontal Assessment Battery (FAB) (Dubois et al., 2000), Trail Making Test (TMT) (completion time and errors for part B) (Godefroy et al., 2010), Rey-Osterreith Complex Figure (Meyers & Meyers, 1995) and Categorical and Phonological Verbal Fluency Tests (Godefroy et al., 2010). They did however identify small significant differences in the Digit Span Forwards and Backwards (Wechsler, 2008) and the number of errors on the TMT-A (Godefroy et al., 2010).

Moreover, a meta-analytic review of 12 papers (n = 497) by Brearly et al. (2017) provided support for the use of video conferencing software to administer a neuropsychological

assessment to a heterogenous clinical sample of patients (such as dementia, mild cognitive impairment, mixed sample of neurological disease and healthy individuals), especially for assessments that rely on verbal responses from participants. However, only two of the studies included in this meta-analytic review drew from a sample of individuals below the age of 65, therefore questioning how transferrable these findings may be to the adult population. Moreover, much of the previous research findings have focused on basic screening tools which may be insensitive to detecting neurocognitive deficits, such as Loh et al. (2007) and Montani et al. (1997) who both administered the MMSE (Folstein et al., 1975), thus emphasising the need for assessing the effectiveness of a broader, more detailed neuropsychological assessment battery.

Rationale and aim

In order to assess whether a virtual administration of a neuropsychological assessment battery provide valid and comparable results to a face-to-face administration of a neuropsychological assessment battery, neuropsychological tests must be administered to a normative sample. Following this, these tests can be administered to a clinical population to establish norms for individuals with distinct pathology. Therefore, the aim of the current research was to administer a battery of neuropsychological assessments to a normative sample, to establish an individual's neuropsychological profile using telecommunication software to establish equivalence of tele-administration of assessments.

Predictions and hypotheses

In light of the previous research findings, a null hypothesis which predicted that there will be no difference between administering a battery of neuropsychological assessments virtually compared with face-to-face administration was adopted.

Methodology

Design

A counterbalanced, within-subject's experimental design was employed to assess the convergent validity of face-to-face and virtual administration of a neuropsychological assessment battery. Participants were assessed twice, once face-to-face, and once using telecommunication software (such as Zoom or Microsoft Teams). Participants were required to complete the assessment on a computer or laptop with a minimum screen size of thirteen inches. The study was counterbalanced as half of the participants firstly completed the neuropsychological assessment battery face-to-face, and the other half being administered the neuropsychological assessment battery using telecommunication software first. This was to help reduce practice effects.

Recruitment and procedure

Eligible participants were recruited between January and April 2022. The project was advertised using social media and posters advertised at the School of Psychology at the University of Birmingham (see appendix 1). Eligible participants contacted the researcher who then arranged a time to discuss the participant information sheet (see appendix 2) with the prospective participant, and to answer any questions they may have. If the prospective participant was happy to take part, then consent was taken (two wet ink consent forms were completed, one for the participant and one to retain in the site file) (see appendix 3). Participants were then randomly allocated to receive either the face-to-face or the virtual neuropsychological assessment battery first and a suitable time was agreed upon to complete the assessments at the School of Psychology, University of Birmingham. Paper was also given to the participant for their virtual assessment and retrieved following the assessment. Participants were not incentivised for taking part.

The assessment battery took around two hours to complete, after which a suitable time to complete the second part of the assessment (either virtual or face-to-face depending on the mode of their first assessment) was agreed upon. Following completion of the assessment, personal data was stored separately to the research data. Participants were allocated an alpha numeric code for identification and data was kept for 10 years in line with the University of Birmingham's policy. Participants were made aware that they can choose to withdraw their data from the study up until the analysis, and their data will be destroyed, and their contact details removed from the database if requested.

Inclusion/exclusion criteria

Based on the normative data available for some of the tests in the assessment battery, adults aged between 18 and 89 wereeligible to take part in the study. Additionally, those without a diagnosis of a neurological condition or learning disability who were English speakers (to a sufficient standard that it would not invalidate the standard administration of the test) and able to give informed consent were eligible to take part in the study. Participants also needed access to the internet and to be able to use telecommunication software (either Zoom, Skype or Microsoft Teams) and access to an appropriate screen size.

Materials

The neuropsychological assessment battery was chosen to provide an overview of an individual's cognitive functioning.

Motivational and effort testing

Participants performance validity was assessed using the stand alone Test of Memory Malingering (TOMM), trial one (Tombaugh, 1996) (see appendix 4). The TOMM has been shown to be both a valid and reliable measure of effort (Sollman & Berry, 2011). In addition to the stand-alone performance validity test (PVT), there were embedded PVT's within some

of the neuropsychological assessments. The Reliable Digit Span score of the Wechsler Adult Intelligence Scale 4th Edition (WAIS-IV) (Wechsler, 2008) (see appendix 5) Digit Span was used, and has been found to have 93% specificity when set at 6 (Babikian et al., 2006). Moreover, the recognition condition of the Wechsler Memory Scale 4th Edition (WMS-IV) (Wechsler, 2009) Logical Memory test and Visual Reproduction test (see appendix 6) were used to assess effort, and have been found to have high sensitivity and specificity (Bouman et al., 2016; Soble et al., 2019).

General intellectual functioning

Premorbid Functioning

The Test of Premorbid Functioning - United Kingdom ($TOPF^{UK}$) (Wechsler, 2008) (see appendix 7) was used to assess premorbid functioning. The $TOPF^{UK}$ consists of seventy words that increase in unfamiliarity and irregularity. This test helps to determine an approximate premorbid level of functioning. The $TOPF^{UK}$ manual not only reports excellent internal consistency (r = .92 - .99) and test-retest reliability (r = .89 - .95), the $TOPF^{UK}$ also correlates with the other WAIS-IV measures of general intellectual functioning (r = .70), and verbal intelligence (r = .75) (Holdnack & Drozdick, 2009).

Current Intellectual Function

The WAIS-IV (Wechsler, 2008) Information (see appendix 8) and Matrix Reasoning (see appendix 9) sub-tests were used to assess current intellectual functioning. The information sub-test consists of twenty-six general knowledge questions which increase in difficulty as the test progresses. The Matrix Reasoning test is a twenty-six item, untimed, sub-test which assesses perceptual organisation, visual processing, and abstract, and spatial perception. Participants were presented with an incomplete matrix and were required to select the appropriate response to complete the matrix.

Memory and Attention functioning

Participants were administered two subtests from the WMS-IV (Wechsler, 2009): a test of verbal episodic memory (Logical Memory) (see appendix 10) and a test of recall for non-verbal visual stimuli (Visual Reproduction) (see appendix 11). For the Logical Memory sub-test, participants listened to two short stories and were asked to recount all they could remember from the stories. This was repeated 20 – 30 minutes later. The Visual Reproduction task required participants to reproduce designs immediately after seeing them, and then 20 – 30 minutes after presentation.

The Digit Span sub-test of the WAIS-IV (Wechsler, 2008) was used to assess attention, concentration, verbal and working memory. Participants were verbally presented with a string of number which increases in length after a correct response. Participants were firstly required to repeat the number, then repeat them in reverse, and finally repeat the numbers in numerical order.

The Rey-Osterreith Complex Figure task (Osterrieth, 1944; Rey, 1941) (see appendix 12) is a measure of perceptual-organizational and constructional ability, and visual-perceptual memory. Participants were asked to copy the complex figure, and then reproduce the figure from memory three minutes (immediate recall) and 30 minutes (delayed recall) after presentation.

Executive functioning

The Oral Trail Making Test (Ricker & Axelrod, 1994) (see appendix 13) assesses attention, tracking and maintenance of cognitive set-shifting. Participants were first asked to count from one to twenty-five as quickly as possible, but without making any errors. For the second condition, participants were required to alternate between numbers and letters as quickly as possible without making any errors.

Two sub-tests of the Delis-Kaplan Executive Function System test battery (Delis et al., 2001) were administered. The Verbal Fluency test (see appendix 14) was used as it evaluates the spontaneous cognitive initiation, set-shifting and cognitive flexibility while under restricted search conditions (firstly generate words beginning with a specific letter, then within a specific category, and finally producing words while switching between two categories). The Colour-Word Interference test (see appendix 15) was used to measure the participants ability to maintain a goal and suppress an habitual response. Participants were firstly required to name colour patches on a page, and then read words aloud. Then, participants are presented with words printed in a different coloured ink to the word and were required name the colour of the word and not read the word aloud (therefore suppressing the habitual response to read the word). Finally, participants were required to switch between rules requiring participants to either read the word aloud or name the colour.

The Hayling Sentence Completion test (see appendix 16) and the Brixton Spatial Anticipation Tests (Burgess & Shallice, 1997) (see appendix 17) were used to evaluate initiation speed and response suppression as well as rule learning and cognitive flexibility. In the Hayling Sentence Completion test, participants were asked to complete several sentences, firstly providing a word that is congruent with the sentence, and then providing an unconnected word (and therefore inhibit an automatic response), as quickly as possible. The Brixton Spatial Anticipation test requires participants to detect a rule to predict the positioning of blue circle to one of ten positions, which then changes without warning (therefore requiring participants to identify the new rule).

Information processing speed

The Oral Symbol Modalities Test (Smith, 1973) (see appendix 18) was used to assess information processing speed, divided attention, visual scanning, and tracking. Participants

were presented with nine symbols with a corresponding number, presented in a legend above the test items. Participants were then required to pair the number that corresponded to a symbol as quickly as possible within 90 seconds.

Language functioning

The full version of the Boston Naming Test (Kaplan et al., 1983) (see appendix 19) was used to asses language function, and consisted of 60 images, presented in order of difficulty. Participants were required to identify the image within twenty seconds, however, if the image was incorrectly identified, then a stimulus cue was offered, followed by a phonemic cue (if the participant is still incorrect).

In addition to the neuropsychological assessment battery, participants were also asked to provide demographic information consisting of their date of birth, ethnicity, gender, and number of years in education.

Ethical approval and issues

Ethical approval was sought and received from the University of Birmingham Research Ethics Committee (ERN_21-1412) (see appendix 20). A comprehensive risk assessment with the School of Psychology was also carried out, assessing the additional risk of face-to-face assessments during the COVID-19 pandemic. Approval from the School of Psychology's Risk Assessment Committee was also sought and received (RA SOPHS_21_100_CJ) (see appendix 21).

Statistical analysis

A database using Statistical Package for the Social Sciences (SPSS v.22) was created to manage the statistical analyses. All analyses carried out were of one-tailed significance unless otherwise stated and the Alpha level was set at .05. Descriptive statistics were explored, and an analysis

of covariance model was created for each variable (face-to-face and its corresponding virtual variable), using the raw scores from each assessment, controlling for age, the date of the first assessment and days between the assessments. These were entered as covariates to control for these variables in the analysis to determine if there were any differences between administering the neuropsychological assessment battery online to virtual administration.

Results

Demographics and descriptive statistics

Twenty-eight healthy participants were recruited and assessed face to face and virtually with a battery of commonly used psychometric tests. The mode of administration (i.e. face-to-face or virtual administration) was counterbalanced across participants. The order of the psychometric tests within a mode of administration was the same for all participants.

The demographic and descriptive statistics of participants can be found in *Table 11*.

Table 11 demographics and descriptive statistics

		Male			Female		
		Count	**Mean**	**Standard Deviation**	**Count**	**Mean**	**Standard Deviation**
	White	11			14		
	Asian	0			2		
Ethnicity[1]	**Mixed**	0			1		
Age[2]		11	41.27_a	15.15	17	39.12_a	12.14
Years in education[2]		11	14.09_a	2.88	17	15.06_a	2.30

[1] Chi square test of sex by ethnicity = X^2 -= 2.174, p = 0.337, exact p = 0.505

[2] Values in the same row not sharing the same subscript are significantly different at p< .05 in the two-sided test of equality for column means. Cells with no subscript are not included in the test. Tests assume equal variances.

Time differences between face to face and virtual administration

The difference in the length of time between first and second assessment is shown in *Table 12*.

Table 12 significant difference in the length of time between first and second assessment

	First Assessment					
	Face to face			**Virtual**		
	Mean	**Standard Deviation**	**Count**	**Mean**	**Standard Deviation**	**Count**
Days between assessments	24	18	14	50	26	14

Note: Values in the same row and sub-table not sharing the same subscript are significantly different at p < .05 in the two-sided test of equality for column means. Cells with no subscript are not included in the test. Tests assume equal variances.

As there was a significant difference in the length of time between first and second assessment (t = -3.035, p = 0.005) in those participants that received the face-to-face assessment first compared to those who firstly received the virtual assessment, the length of time between first and second assessment will be used as a covariate in the subsequent analysis of discrepancies between face-to-face and virtual test administration. This covariate is included because it is plausible that practice effects may vary as a function of the length of time since the initial test administration.

Discrepancies between face-to-face and virtual test administration

The discrepancy between face-to-face and virtual test administration was assessed using a four-way analysis of covariance (ANCOVA). The within-subject factor was the mode of administration (face-to-face versus virtual administration) and the between subjects' factor was the counterbalancing of the initial mode of administration.

Two covariates were also included. As previously noted, the average number of days between first and second test administration was significantly different depending on whether the first administration of the test was face-to-face or virtual. As the length of time between administrations could influence practice effects then the length in time (in days) between test administrations was included as a covariate. The second covariate was the age (in years) of the participant. This was included as some of the psychometric tasks show an ageing profile it is possible that raw scores may be influenced by the age of the participant.

The age and administration delay covariates appearing in the ANCOVA model were evaluated at their average values for the participants undertaking testing (age = 39.96 years and administration delay = 36.71 days).

Motivational and effort testing

In order to assess a participants effort level, one stand alone and three embedded measures of test validity were used. Participants were excluded if they scored below the established cut off score of 42 on the TOMM (Trial 1) (Martin et al., 2020). *Table 13* shows the cut off score for the embedded effort measures.

Table 13: Cut off scores for the embedded effort measures

Test	Cut off score	Reference
Logical memory (recognition)	15	Holdnack et al. (2013)
Visual reproduction (recognition)	3	Holdnack et al. (2013)
Reliable Digit Span	7	Holdnack et al. (2013)

Based on this criterion, none of the participants included in this study scored below 42 on the TOMM, therefore none of the participants data was removed from the analysis.

In order to assess which mode of administration resulted in differences in performance on the Test of Memory Malingering (trial one) a four-way analysis of covariance was constructed as indicated above (see *Table 14*).

Table 14: Dedicated measures of test validity,

	Face to Face administration		**Virtual administration**		**Age cohort effect**	**First Administration effect**	**Days between Administration effect**	**Administration effect**
	Mean	**SD**	**Mean**	**SD**	**F (*p*)**	**F (*p*)**	**F (*p*)**	**F (*p*)**
TOMM Trial 1	48.179	0.457	48.607	0.313	0.073 (0.789)	0.555 (0.464)	0.007 (0.934)	< 0.001 (0.993)

There was no significant difference between the face-to-face and virtual administration of the Test of Memory Malingering (F = < 0.001, p = 0.993). Similarly, order of administration, the delay between administration and the age of participants did not effect performance on this task.

A four-way ANCOVA was also constructed to assess performance on embedded measures of test validity differed by mode of administration (see *Table 15*). The tests were the recognition task of the Logical Memory and Visual Reproduction tasks within the WMS – IV (Wechsler, 2009) and the Reliable Digit Span (Holdnack & Drozdick, 2009) calculated from the Digit Span test within the WAIS – IV (Wechsler, 2008).

Table 15: Embedded measures of test validity

	Face to Face administration		Virtual administration		cohort Age effect	First Administration effect	Days between Administration effect	Administration effect
	Mean	SD	Mean	SD	F (*p*)	F (*p*)	F (*p*)	F (*p*)
Logical memory (recognition)	25	3.63	26.18	2.97	0.023 (0.880)	1.106 (0.304)	1.190 (0.286)	0.435 (0.516)
Visual reproduction (recognition)	6.32	0.941	6.64	0.488	1.651 (0.211)	0.174 (0.680)	1.842 (0.187)	0.003 (0.960)
Reliable Digit Span	9.14	2.050	9.43	2.348	0.743 (0.397)	0.015 (0.903)	0.126 (0.726)	0.110 (0.744)

There was no significant difference between the face-to-face and virtual administration on any of the embedded measure of test validity (Logical Memory, F = 0.435, p = 0.516; Visual Reproduction, F = 0.003, p = 0.960; Reliable Digit Span, F = 0.110, p = 0.744) (Holdnack & Drozdick, 2009; Wechsler, 2009). Similarly, order of administration, the delay between administration and the age of participants did not effect performance on this task.

General intellectual functioning

Premorbid Functioning

The Test of Premorbid Functioning - United Kingdom (TOPF^{UK}) (Wechsler, 2008) was used to assess premorbid functioning. *Table 16* shows a four-way ANCOVA, which was constructed to assess differences in mode of administration.

Table 16: Test of Premorbid Functioning

	Face to Face administration		Virtual administration		Age cohort effect	First Administration effect	Days between Administration effect	Administration effect
	Mean	SD	Mean	SD	F (*p*)	F (*p*)	F (*p*)	F (*p*)
TOPF^{UK}	42.61	13.11	43.51	11.70	0.937 (0.343)	4.697 (0.040)	0.323 (0.575)	0.562 (0.461)

There was no significant difference between the face-to-face and virtual administration of TOPF^{UK} (Wechsler, 2008) scores (F = 0.562, p = 0.461). Participants age and the delay between test administration did not have a significant effect on test performance, however order of presentation did have a significant effect, whereby face to face administration scores appear lower in those who received the face-to-face administration first, whereas virtual administration scores appeared higher in those who received the virtual administration first compared to those who received virtual administration second.

Current Intellectual Function

Two sub-tests from the WAIS-IV (Information and Matrix Reasoning) (Wechsler, 2008) were used to assess current intellectual functioning. Raw scores were entered into a four-way analysis of covariance, as indicated above. The results are presented in *Table 17*.

Table 17: Current intellectual functioning assessed using the WAIS-IV Information and Matrix Reasoning subscales

	Face to Face administration		Virtual administration		Age cohort effect	First Administration effect	Days between Administration effect	Administration effect
	Mean	SD	Mean	SD	F (*p*)	F (*p*)	F (*p*)	F (*p*)
Information	17	4.56	17	4.51	1.587 (0.220)	0.005 (0.947)	1.820 (0.190)	0.157 (0.695)
Matrix reasoning	18.11	4.67	17.36	4.066	1.477 (0.236)	0.023 (0.880)	0.869 (0.361)	2.163 (0.154)

There was no significant difference between the face-to-face and virtual administration of either test of the WAIS-IV (Wechsler, 2008) (Information, F = 0.157, p = 0.695; Matrix Reasoning, F = 2.163, p = 0.154). Moreover, order of administration, the delay between administration and the age of participants did not affect performance on this task.

Memory and Attention functioning

In order to explore any differences in the mode of administration in memory and attention tasks, two scales from the WMS-IV (Wechsler, 2009), one from the WAIS-IV (Wechsler, 2008) and the Rey-Osterreith Complex Figure (Osterrieth, 1944; Rey, 1941) were administered. A four-way analysis of covariance was constructed for each of these tests, as indicated above (*Table 18*).

Table 18: Memory and attention functioning, assessed using the WMS-IV Logical Memory, Visual Reproduction, the WAIS-IV Digit span, and the Rey Osterreith Complex Figure.

	Face to Face administration		Virtual administration		Age cohort effect	First Administration effect	Days between Administration effect	Administration effect
	Mean	SD	Mean	SD	F (*p*)	F (*p*)	F (*p*)	F (*p*)
WMS IV Logical Memory								
Immediate	29.79	8.006	32.50	8.677	0.176 (0.678)	2.115 (0.159)	2.001 (0.170)	1.846 (0.187)
Delayed	27.36	8.385	28.36	8.341	0.305 (0.586)	23.009 (<.001)	1.955 (0.175)	1.540 (0.227)
WMS-IV Visual recognition								
Immediate	38.68	4.730	38.61	4.856	0.240 (0.628)	7.366 (0.012)	0.167 (0.686)	0.055 (0.817)
Delayed	35.04	7.234	35.43	5.534	0.896 (0.353)	9.720 (0.005)	0.120 (0.732)	0.573 (0.457)
WAIS-IV Digit Span								
Forwards	10.18	2.568	10.61	2.601	0.468 (0.500)	4.203 (0.051)	2.845 (0.105)	0.757 (0.393)
Backwards	7.00	2.108	6.50	2.301	1.758 (0.179)	0.349 (0.560)	1.244 (0.276)	0.662 (0.424)
Sequencing	8.50	1.202	8.57	1.752	1.003 (0.327)	1.569 (0.222)	3.419 (0.077)	0.002 (0.965)
Total	25.68	5.004	25.68	5.538	2.983 (0.097)	0.000 (0.995)	0.681 (0.417)	0.906 (0.351)
Rey-Osterreith Complex Figure								
Copy	34.32	2.310	34.18	2.881	2.114 (0.159)	0.115 (0.737)	0.277 (0.604)	0.832 (0.371)
Immediate	23.089	7.789	23.125	7.069	0.190 (0.666)	3.459 (0.075)	0.022 (0.883)	0.067 (0.798)
Delayed	21.946	7.754	22.679	7.596	0.190 (0.667)	9.713 (0.005)	0.161 (0.691)	0.000 (0.992)

There was no significant difference between the face-to-face and virtual administration of any tests used to assess memory and attention. Participants age and the delay between test administration did not have a significant effect on test performance, however order of presentation did have a significant effect on the Delayed Logical Memory task and the Delayed

Visual Recognition task (Wechsler, 2009), whereby scores were higher when participants repeated the assessment, and on the Delayed Rey-Osterreith Complex Figure (Osterrieth, 1944; Rey, 1941), with scores being higher when participants repeated the assessment. There was also a trending significant order effect on the WAIS-IV Digit Span (Wechsler, 2008).

Executive functioning

The Verbal Fluency and Colour Word Interference Tests of the DKEFS test battery (Delis et al., 2001), the Oral Trail Making Test (Ricker & Axelrod, 1994) and the Hayling and Brixton Tests (Burgess & Shallice, 1997) were used to assess executive functioning. In order to assess which mode of administration resulted in differences in performance on these tests, a four-way analysis of covariance was constructed as indicated above (*Table 19*).

Table 19: Executive functioning, assessed using the Hayling and Brixton tests and the Colour Word Interference and Verbal Fluency tests of the DKEFS test battery.

	Face to Face administration		**Virtual administration**		**Age cohort effect**	**First Administration effect**	**Days between Administration effect**	**Administration effect**
	Mean	**SD**	**Mean**	**SD**	**F (*p*)**	**F (*p*)**	**F (*p*)**	**F (*p*)**
Hayling Sentence Completion (time)	46.21	32.004	59.18	32.740	0.004 (0.949)	2.556 (0.123)	5.436 (0.028)	0.167 (0.686)
Brixton Spatial Anticipation Test (errors)	16.86	9.236	15.86	6.311	3.136 (0.089)	5.626 (0.026)	0.050 (0.825)	1.543 (0.226)
DKEFS Colour-Word Interference test								
Uncorrected colour naming	0.00	0.00	0.14	0.448	5.619 (0.026)	2.876 (0.103)	0.533 (0.472)	2.723 (0.112)
Corrected colour naming	0.32	0.25	0.25	0.441	5.559 (0.027)	0.283 (0.599)	0.733 (0.401)	4.912 (0.036)
Colour naming time	28.07	4.906	28.32	6.377	0.840 (0.368)	11.406 (0.002)	4.225 (0.051)	0.046 (0.833)
Word reading Uncorrected	0.00	0.000	0.04	0.189	0.296 (0.592)	0.188 (0.669)	7.029 (0.014)	1.500 (0.233)
Word reading Corrected	0.07	0.262	0.11	0.315	0.038 (0.847)	0.057 (0.814)	0.230 (0.636)	0.048 (0.828)
Word reading Time	21.96	5.847	21.54	4.887	0.060 (0.809)	1.679 (0.207)	0.179 (0.676)	0.011 (0.916)
Inhibition Uncorrected	0.36	1.062	0.25	0.645	0.175 (0.680)	0.757 (0.393)	2.719 (0.112)	0.207 (0.653)
Inhibition Corrected	0.68	1.056	0.75	1.143	2.131 (0.157)	1.891 (0.182)	2.010 (0.169)	2.475 (0.129)
Inhibition Time	52.18	15.435	52.57	14.992	0.323 (0.575)	0.589 (0.450)	8.020 (0.009)	2.198 (0.151)

	Face to Face administration		Virtual administration		Age cohort effect	First Administration effect	Days between Administration effect	Administration effect
	Mean	SD	Mean	SD	F (*p*)	F (*p*)	F (*p*)	F (*p*)
Inhibition/switching Uncorrected	0.54	1.036	0.50	1.00	6.258 (0.020)	2.436 (0.132)	0.182 (0.674)	3.930 (0.059)
Inhibition/switching Corrected	0.68	1.020	1.14	1.484	0.006 (0.938)	0.008 (0.930)	0.400 (0.533)	0.011 (0.919)
Inhibition/switching Time	64.50	31.233	58.79	13.362	0.835 (0.370)	2.311 (0.142)	3.085 (0.092)	0.078 (0.782)
DKEFS Verbal Fluency test								
Letters	43.82	10.951	44.11	12.764	3.980 (0.058)	8.659 (0.007)	0.982 (0.332)	3.287 (0.082)
Category	50.75	7.457	51.75	8.081	0.015 (0.905)	2.607 (0.119)	0.077 (0.784)	0.002 (0.965)
Switching	16.32	2.932	16.96	2.795	0.562 (0.461)	11.650 (0.002)	1.584 (0.220)	0.567 (0.459)
Oral Trail Making Test time	36.21	18.17	40.00	21.15	0.067 (0.797)	2.211 (0.150)	0.200 (0.659)	.0095 (0.760)
Oral Trail Making Test Set Loss errors	0.21	0.499	0.43	0.742	0.457 (0.506)	1.387 (0.251)	1.238 (0.277)	1.444 (0.241)
Oral Trail Making Test sequential errors	0.82	1.467	0.71	0.976	0.131 (0.720)	1.645 (0.212)	0.246 (0.625)	0.290 (0.595)

There was a significant difference between face to face and virtual administration on the colour naming task (Delis et al., 2001). Participants performed significantly better in correcting an error made in the colour naming task when they were administered the virtual test than they did face to face. There was also an age cohort effect on the corrected colour naming task, uncorrected colour naming task and the uncorrected inhibition/switching task (Delis et al., 2001). Moreover, there was also a trending significant difference between the inhibition/switching uncorrected score, with participants making fewer errors on the virtual administration of the task compared to face-to-face administration. There was also a significant effect of delay between the follow up assessment on the word reading uncorrected score and a trending significant effect of delay between the follow up assessment on the time to complete the colour naming task. Order of presentation also had a significant effect on the colour naming time, with individuals performing better when repeating the assessment.

A trending significant difference was found in the letter fluency condition of the DKEFS verbal fluency test (Delis et al., 2001), with participants generating more words beginning with a specified letter when administered the task virtually, compared to face-to-face administration. Participants age and the delay between test administration did not have a significant effect on test performance, however order of presentation did have a significant effect on the letters and switching condition of the DKEFS Verbal Fluency task (Delis et al., 2001), whereby scores were higher when participants repeated the assessment.

There was a significant effect on order of presentation on the Brixton Spatial Anticipation task (Burgess & Shallice, 1997), with participants performing better when repeating the assessment, and a trending age cohort effect. There was however no overall effect of mode of administration. On the Hayling Sentence Completion task (Burgess & Shallice, 1997), there was a significant effect of delay between the follow up assessment, but there was no significant difference between mode of administration.

There was no significant differences between virtual and face-to-face administration of the Oral Trail Making Test (Ricker & Axelrod, 1994), and there was no age cohort or administration effects on these assessments.

Information processing speed

The Oral Symbol Modalities Test (Smith, 1973) was used to assess information processing speed. In order to assess which mode of administration resulted in differences in performance on this test, a four-way analysis of covariance was constructed as indicated above (*Table 20*).

Table 20: Processing speed assessed using the Oral Symbol Modalities Test.

	Face to Face administration		Virtual administration		Age cohort effect	First Administration effect	Days between Administration effect	Administration effect
	Mean	SD	Mean	SD	F (*p*)	F (*p*)	F (*p*)	F (*p*)
Oral Symbols Modalities Test	56.07	15.639	57.79	13.248	1.323 (0.261)	7.670 (0.011)	2.350 (0.138)	1.445 (0.241)

Performance on the Oral Symbol Modalities Test (Smith, 1973) did not significantly differ between face-to-face and virtual administration (F = 1.445, p = 0.241). Order of presentation had a significant effect on test performance, with individuals performing better when repeating the assessment. The delay between administration and the age of participants did not affect performance on this task.

Language functioning

To assess which mode of administration resulted in differences in performance on the Boston Naming Test (Kaplan et al., 1983), a four-way analysis of covariance was constructed as indicated above (*Table 21*).

Table 21: Language function test.

	Face to Face administration		Virtual administration		Age cohort effect	First Administration effect	Days between Administration effect	Administration effect
	Mean	SD	Mean	SD	F (p)	F (p)	F (p)	F (p)
Boston Naming Test	55.61	3.247	55.82	3.486	1.088 (0.307)	7.440 (0.012)	1.617 (0.216)	0.015 (0.903)

Scores on the Boston Naming Test (Kaplan et al., 1983) did not significantly differ between the face-to-face and virtual administration (F = 0.015, p = 0.903). Order of presentation had a significant effect on test performance, with individuals performing better when repeating the assessment. The delay between administration and the age of participants also did not affect performance on this task.

Discussion

Summary of findings

The aim of this study was to explore if there are differences between face-to-face and virtual administration of a battery of neuropsychological assessments. The results from this small-scale study identified significant differences in the DKEFS (Delis et al., 2001) corrected Colour Naming task, with those being administered the task virtually performing significantly better than when completing the task face-to-face. There was also a trending significant difference in mode of administration in the DKEFS (Delis et al., 2001) inhibition/switching uncorrected score, again with participants performing better when being administered the test virtually compared to face-to-face administration. Moreover, there was a trending significant difference in the number of words generated beginning with a specific letter in the DKEFS verbal fluency task (Delis et al., 2001), with participants performing better in the virtual administration condition compared to face-to-face administration, although this was not a significant association.

There were no significant differences in mode of administration for the tests assessing motivation and effort and there were no significant differences between virtual and face-to-face administration of the WMS-IV (Wechsler, 2009) tests (Logical Memory and Visual Reproduction), the WAIS-IV (Wechsler, 2008) (Test of Premorbid Functioning, Information, Matrix Reasoning and Digit Span) or in any of the three conditions of the Rey-Osterreith Complex Figure (Osterrieth, 1944; Rey, 1941). Regarding the tests of executive functioning, there was no significant difference in mode of administration for the Hayling and Brixton tests (Burgess & Shallice, 1997) or the Oral Trail Making Test (Ricker & Axelrod, 1994). Participants also did not differ in their performance on the Oral Symbol Modalities Test and

Boston Naming Test (Kaplan et al., 1983) when administered face-to-face compared with virtual administration.

The absence of a significant relationship between many of the neuropsychological assessments used in this study support previous research exploring differences in performance in virtual and face-to-face neuropsychological assessments. Munro Cullum and Grosch (2013) found that there was no statistically significant difference in performance of the DKEFS Category Fluency (Delis et al., 2001), Boston Naming Test (Kaplan et al., 1983) and Digit Span (Wechsler, 2008) when administered virtually compared with face-to-face administration, which was supported by the present research. Moreover, there was a lack of significant associations between mode of assessment and performance on the Trail Making Test (Delis et al., 2001), Rey-Osterreith Complex Figure (Osterrieth, 1944; Rey, 1941) and the DKEFS Category Fluency test (Delis et al., 2001), which supports the findings by Gnassounou et al. (2021). Finally, a study by Hildebrand et al. (2004) found no significant difference between virtual and face-to-face administration of the WAIS-IV Matrix Reasoning, in a sample of older adults.

A meta-analytic review of twelve studies (n = 497) by Brearly et al. (2017) found that performance on the verbally mediated tasks (such as the digit span, verbal fluency and list learning) did not significantly differ when administered virtually or face-to-face. The present findings partially support these conclusions given the lack of significant differences found in performance on the Digit Span when administered virtually and face-to-face, however the present study identified a trending significant association between mode of neuropsychological assessment and performance on the Verbal Fluency task. However, many of the studies in the meta-analytic review by Brearly et al. (2017) re-assessed participants on the same day, which differs from the present study, therefore the association identified by Brearly et al. (2017) may be due to practice effects.

There is a scarcity of research examining the validity of virtual administration of the DKEFS Colour-Word Interference test (Delis et al., 2001), which provides an avenue for future research. Performance on other virtually administered timed executive functioning assessments, which require monitoring of performance, speed and accuracy have been found to differ compared with face-to-face administration, which may support the findings from this study. A study of fifty-five healthy controls compared with twenty-nine participants with Mild Cognitive Impairment or Dementia by Wadsworth et al. (2016) found significant differences in performance on the Trail Making Task when administered virtually compared with face-to-face administration. However, the present study did not find differences in performance on the oral version of the Trail Making Test (Ricker & Axelrod, 1994). Moreover, despite there being statistically significant differences between performance on the virtually administered DKEFS Colour-Word Interference test (Delis et al., 2001), these differences may not reflect a clinically significant difference and may be influenced by a limited sample size. One explanation for the superior performance on the DKEFS Colour-Word Interference task (Delis et al., 2001) when administered virtually compared with face-to-face administration may be that a smaller screen reduced the spaces between the colours and words, which may explain the differences in performance.

Participants age also significantly impacted performance on the DKEFS Colour-Word Interference task (Delis et al., 2001), with participants making fewer errors when administered the colour naming task and inhibition/switching task virtually compared to face-to-face. Moreover there was a non-significant, trending significant association between age and performance when participants were asked to generate words beginning with a specific letter (letters test of the DKEFS Verbal Fluency test (Delis et al., 2001). This suggests that there may be an age cohort effect on tests of executive function, indicating that performance on the repeated neuropsychological assessment battery may have been influenced by the participant's

age, with younger participants improving in the repeated executive functioning tasks compared to older participants. These findings support previous research exploring executive functioning. A study of three hundred and fifty healthy participants aged between ten and eighty-six by Ferguson et al. (2021) found that performance on tests of executive functioning (such as the Stroop task), was significantly associated with age, with individuals aged between ten and thirty-six showing an improved inhibitory control compared to those aged between thirty-six and eighty-six who showed a decline in inhibitory control.

Clinical implications

The findings from this study indicate that, with the exception of certain tests of executive functioning, performance on a battery of neuropsychological assessments administered virtually was comparable to performance when administered face-to-face for a normative population. One implication of this is that a valid neuropsychological assessment can be carried out virtually therefore, removing the necessity for patients to attend a face-to-face clinic for a neuropsychological assessment. However, consideration must be given to test selection, given the difference in test performance on the DKEFS Colour-Word Interference test (Delis et al., 2001) when administered virtually compared to face-to-face administration. Although there is a caveat to virtual neuropsychological assessments. Conducting an assessment using telecommunication software may impact on a clinicians ability to observe and document behaviour displayed during an assessment, which may be exacerbated when assessing an individual from a culturally diverse background (Bilder et al., 2020). However, the present findings indicate that, where it may not be possible to conduct a face-to-face neuropsychological assessment, that valid results are yielded in most assessments that made up the neuropsychological battery when administered virtually.

One important caveat to the findings is that despite performance on formal neuropsychological testing being comparable when administered virtually compared with face-to-face administration, it is important to note that conducting assessments virtually may add the benefit of convenience, but at the expense of a strong therapeutic alliance, which forms the bedrock of Clinical Psychology as a profession and is essential in psychotherapeutic work and may be at risk when working exclusively with patients virtually (Cataldo et al., 2021). Therefore shifting entirely to a model of virtual assessments and therapy, devoid of human contact and face-to-face interaction is wholly incongruous with the values and philosophical underpinnings of the profession of Clinical Psychology.

Limitations

One potential limitation with the current study is the limited sample size. The study recruited twenty-eight participants; therefore statistical analyses may be underpowered for statistical analysis. However, to overcome this limitation, a within-subject's design was employed. Another limitation with the current research is the lack of acceptability measure. Although not systematically or routinely collected, many participants offered an account of their experiences after the assessment, and with some reporting that they believed their performance to be better when the neuropsychological assessment was administered virtually compared with face-to-face administration, while others preferred face-to-face administration. Therefore a systematic recording of the participants experiences and mood measures may have enriched the data and contextualised some of the findings.

Future directions

The findings suggest that there may be differences in mode of administration and performance on some neuropsychological assessments, specifically tests assessing executive functioning (such as the DKEFS Colour-Word Interference test (Delis et al., 2001). Therefore future

research should focus on understanding these differences, with a larger sample size, and discerning if these differences are clinically meaningful. Moreover, future research should focus on examining the validity of neuropsychological assessments being administered virtually, with specific patient groups (such as stroke patients).

The rapid acceleration of teleneuropsychology since the beginning of the COVID-19 pandemic has augmented research in this emerging area. While future research should continue to validate virtual administration of existing neuropsychological assessments, consideration must be given for a paradigm shift in clinical neuropsychological assessment, which moves away from traditional face-to-face assessment using a pen and paper, to a more refined and nuanced neuropsychological assessment battery using information technology. One criticism that has been levelled of clinical neuropsychology is that the neuropsychological assessment relies heavily on outdated methods and is labour intensive (both in terms of data collection, but also in terms of analysis of each assessment) (Miller & Barr, 2017), which may be an inefficient use of time and is open to human error (Collins & Riley, 2016). A neuropsychological assessment battery specifically developed using information technology may provide a more accurate and sensitive recording on some of the tasks assessing a patient's speed and reduce the time required to analyse a patient's assessment, as well as eliminating the chance of errors in data entry.

Future research should focus on the acceptability of administering a neuropsychological assessment battery virtually, and perhaps consider utilising a qualitative design to explore patients' experience and preference for mode of administration. This is central to future research in this area given that acceptability of which mode of administration is preferred by the patient is crucial to maintain high standards of patient care.

Finally, concerns surrounding the use of neuropsychological assessments virtually and the security of test material and recording of materials, particularly in a medico-legal setting, need to be reconciled before widespread virtual use. The use of test materials in a setting that cannot be controlled (such as virtually) may compromise the security and integrity of the testing material. Moreover, some publishers of testing materials stipulate that the neuropsychological test should be conducted in an office setting with a technician present to prevent the recording of the material (such as Green's publishing, who have the publishing rights to tests such as the Word Memory Test (Green, 2003) and the Memory Complaints Inventory (Green, 2004).

Conclusion

The findings from this study indicate that, with the exception of some tests of executive functioning, a virtually administered battery of neuropsychological assessments yields valid and comparable results compared with face-to-face administration. There are however avenues for further research including validation of a virtually administered neuropsychological assessment in certain patient groups (such as stroke), and consideration for a bespoke package of neuropsychological assessment created using information technology.

References for the empirical paper

Agrell, B., & Dehlin, O. (1998). The clock-drawing test. *Age and ageing*, *27*(3), 399-404.

Arciniegas, D. B., Anderson, C. A., & Filley, C. M. (2013). *Behavioral neurology & neuropsychiatry*. Cambridge University Press.

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006, Feb). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *Clin Neuropsychol*, *20*(1), 145-159. https://doi.org/10.1080/13854040590947362

Benedict, R. H. B., Schretlen, D., Groninger, L., & Brandt, J. (1998, 1998/02/01). Hopkins Verbal Learning Test – Revised: Normative Data and Analysis of Inter-Form and Test-Retest Reliability. *The Clinical Neuropsychologist*, *12*(1), 43-55. https://doi.org/10.1076/clin.12.1.43.1726

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. (2020). Inter organizational practice committee recommendations/guidance for teleneuropsychology in response to the COVID-19 pandemic. *Archives of Clinical Neuropsychology*, *35*(6), 647-659.

Bondi, M. W., Salmon, D. P., & Kaszniak, A. W. (2009). The neuropsychology of dementia. In *Neuropsychological assessment of neuropsychiatric and neuromedical disorders, 3rd ed.* (pp. 159-198). Oxford University Press.

Bouman, Z., Hendriks, M. P., Schmand, B. A., Kessels, R. P., & Aldenkamp, A. P. (2016). Indicators of suboptimal performance embedded in the Wechsler Memory Scale–Fourth Edition (WMS–IV). *Journal of Clinical and Experimental Neuropsychology*, *38*(4), 455-466.

Brearly, T. W., Shura, R. D., Martindale, S. L., Lazowski, R. A., Luxton, D. D., Shenal, B. V., & Rowland, J. A. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology Review*, *27*(2), 174-186.

Bunnage, M., Evans, J., Wright, I., Thomas, S., Vargha-Khadem, F., Poz, R., Wilson, C., & Moore, P. (2020). *The British Psychological Society Division of Neuropsychology Professional Standards Unit Guidelines to colleagues on the use of Tele-neuropsychology*.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton tests*. Thames Valley Test Company.

Cataldo, F., Chang, S., Mendoza, A., & Buchanan, G. (2021, Feb 19). A Perspective on Client-Psychologist Relationships in Videoconferencing Psychotherapy: Literature Review. *JMIR Ment Health, 8*(2), e19004. https://doi.org/10.2196/19004

Cicerone, K., Levin, H., Malec, J., Stuss, D., & Whyte, J. (2006). Cognitive rehabilitation interventions for executive function: moving from bench to bedside in patients with traumatic brain injury. . *J Cogn Neurosci . 18*(7), 1212 - 1222.

Coetzer, R. (2020, Jul). First Impressions of Performing Bedside Cognitive Assessment of COVID-19 Inpatients. *J Am Geriatr Soc, 68*(7), 1389-1390. https://doi.org/10.1111/jgs.16561

Collins, F. S., & Riley, W. T. (2016, Nov 23). NIH's transformative opportunities for the behavioral and social sciences. *Sci Transl Med, 8*(366), 366ed314. https://doi.org/10.1126/scitranslmed.aai9374

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

Dikmen, S. S., Machamer, J. E., Winn, H. R., & Temkin, N. R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology, 9*(1).

Donders, J., Oh, Y. I., & J., G. (2015). Self- and Informant Ratings of Executive Functioning After Mild Traumatic Brain Injury. *J Head Trauma Rehabil . 30*(6), E30-E39.

Dubois, B., Slachevsky, A., Litvan, I., & Pillon, B. (2000). The FAB: a frontal assessment battery at bedside. *Neurology, 55*(11), 1621-1626.

Ferguson, H. J., Brunsdon, V. E. A., & Bradford, E. E. F. (2021, 2021/01/14). The developmental trajectories of executive function from adolescence to old age. *Scientific Reports, 11*(1), 1382. https://doi.org/10.1038/s41598-020-80866-1

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Gnassounou, R., Defontaines, B., Denolle, S., Brun, S., Germain, R., Schwartz, D., Schück, S., Michon, A., Belin, C., & Maillet, D. (2021). Comparison of Neuropsychological Assessment by Videoconference and Face to Face. *Journal of the International Neuropsychological Society*, 1-11. https://doi.org/10.1017/S1355617721000679

Godefroy, O., Azouvi, P., Robert, P., Roussel, M., LeGall, D., Meulemans, T., & Group, B. o. t. G. d. R. s. l. E. d. F. E. S. (2010). Dysexecutive syndrome: diagnostic criteria and validation study. *Annals of Neurology, 68*(6), 855-864.

Green, P. (2003). Word Memory Test for Windows: User's manual and program. *Edmonton: Green's Publishing*.

Green, P. (2004). *Memory Complaints Inventory.* Green's Publishing.

Grosch, M. C., Gottlieb, M. C., & Cullum, C. M. (2011, 2011/10/01). Initial Practice Recommendations for Teleneuropsychology. *The Clinical Neuropsychologist, 25*(7), 1119-1133. https://doi.org/10.1080/13854046.2011.609840

Hildebrand, R., Chow, H., Williams, C., Nelson, M., & Wass, P. (2004). Feasibility of neuropsychological testing of older adults via videoconference: implications for assessing the capacity for independent living. *J Telemed Telecare, 10*(3), 130-134. https://doi.org/10.1258/135763304323070751

Hilty, D. M., Nesbitt, T. S., Kuenneth, C. A., Cruz, G. M., & Hales, R. E. (2007). Rural versus suburban primary care needs, utilization, and satisfaction with telepsychiatric consultation. *The Journal of Rural Health, 23*(2), 163-165.

Holdnack, J., & Drozdick, L. (2009). Advanced clinical solutions for WAIS-IV and WMS-IV: Clinical and interpretive manual. *Texas: Pearson*.

Holdnack, J. A., Millis, S., Larrabee, G. J., & Iverson, G. L. (2013). Assessing performance validity with the ACS. In *WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation.* (pp. 331-365). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-386934-0.00007-9

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test.* . Lea & Fibiger

Loh, P.-k., Donaldson, M., Flicker, L., Maher, S., & Goldswain, P. (2007). Development of a telemedicine protocol for the diagnosis of Alzheimer's disease. *Journal of Telemedicine and Telecare, 13*(2), 90-94.

Løvstad, M., Sigurdardottir, S., Andersson, S., Grane , V. A., Moberget , T., Stubberud, J., & Solbakk, A. K. (2016). Behavior Rating Inventory of Executive Function Adult Version in Patients with Neurological and Neuropsychiatric Conditions: Symptom Levels and Relationship to Emotional Distress. *J Int Neuropsychol Soc, 22*(6), 682 - 694.

Mahieux-Laurent, F., Fabre, C., Galbrun, E., Dubrulle, A., & Moroni, C. (2009). groupe de réflexion sur les praxies du CMRR Île-de-France Sud1. *revue neurologique, 165*, 560-567.

Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2020, 2020/01/02). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist, 34*(1), 88-119. https://doi.org/10.1080/13854046.2019.1637027

Mazaux, J. M., Masson, F., Levin, H. S., Alaoui, P., Maurette, P., & Barat, M. (1997). Long-term neuropsychological outcome and loss of social autonomy after traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 78*, 1316 - 1320.

Meyers, J. E., & Meyers, K. R. (1995). Rey Complex Figure Test and Recognition Trial: Professional Manual. *Odessa: Psychological Assessment Resource*.

Miller, J. B., & Barr, W. B. (2017). The technology crisis in neuropsychology. *Archives of Clinical Neuropsychology, 32*(5), 541-554.

Montani, C., Billaud, N., Tyrrell, J., Fluchaire, I., Malterre, C., Lauvernay, N., Couturier, P., & Franco, A. (1997). Psychological impact of a remote psychometric consultation with hospitalized elderly people. *Journal of Telemedicine and Telecare, 3*(3), 140-145.

Munro Cullum, C., & Grosch, M. C. (2013). 14 - Special Considerations in Conducting Neuropsychology Assessment over Videoteleconferencing. In K. Myers & C. L. Turvey (Eds.), *Telemental Health* (pp. 275-293). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-416048-4.00014-2

Norman, S. (2006, 2006/12/01). The use of telemedicine in psychiatry [https://doi.org/10.1111/j.1365-2850.2006.01033.x]. *Journal of Psychiatric and Mental Health Nursing, 13*(6), 771-777. https://doi.org/https://doi.org/10.1111/j.1365-2850.2006.01033.x

Osterrieth, P. A. (1944). Le test de copie d'une figure complexe. *Archives of Psychology*(30), 206–356.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique.(Les problems.). *Archives de psychologie*.

Ricker, J. H., & Axelrod, B. N. (1994). Analysis of an oral paradigm for the Trail Making Test. *Assessment, 1*(1), 47-51.

Sigurdardottir, S., Andelic, N., Wehling, E., Roe, C., Anke, A., Skandsen, T., Holthe, O. O., Jerstad, T., Aslaksen, P. M., & Schanke, A.-K. (2015). Neuropsychological functioning in a national cohort of severe traumatic brain injury: Demographic and acute injury-related predictors. *The Journal of Head Trauma Rehabilitation, 30*(2), E1. https://bora.uib.no/bora-

xmlui/bitstream/1956/17537/4/Sigurdardottir%2bet%2bal.%2b2015.%2bPostprint.%2bNeuropsychological%2bFunctioning%2bin%2ba%2bnational%2bcohort%2bof%2bsevere%2btraumatic%2bbrain%2binjury.pdf

Sinanović, O. (2010). Neuropsychology of acute stroke. *Psychiatria Danubina, 22*(2), 278-281.

Smith, A. (1973). *Symbol digit modalities test*. Western psychological services Los Angeles.

Soble, J. R., Bain, K. M., Bailey, K. C., Kirton, J. W., Marceaux, J. C., Critchfield, E. A., McCoy, K. J. M., & O'Rourke, J. J. F. (2019, Jul-Aug). Evaluating the accuracy of the Wechsler Memory Scale-Fourth Edition (WMS-IV) logical memory embedded validity index for detecting invalid test performance. *Appl Neuropsychol Adult, 26*(4), 311-318. https://doi.org/10.1080/23279095.2017.1418744

Sollman, M. J., & Berry, D. T. (2011, Dec). Detection of inadequate effort on neuropsychological testing: a meta-analytic update and extension. *Arch Clin Neuropsychol, 26*(8), 774-789. https://doi.org/10.1093/arclin/acr066

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. Multi Health Systems.

Van der Linden, M., Coyette, F., Poitrenaud, J., Kalafat, M., Calicis, F., Wyns, C., & Adam, S. (2004). II. L'épreuve de rappel libre/rappel indicé à 16 items (RL/RI-16).

Wadsworth, H. E., Galusha-Glasscock, J. M., Womack, K. B., Quiceno, M., Weiner, M. F., Hynan, L. S., Shore, J., & Cullum, C. M. (2016, Aug). Remote Neuropsychological Assessment in Rural American Indians with and without Cognitive Impairment. *Arch Clin Neuropsychol, 31*(5), 420-425. https://doi.org/10.1093/arclin/acw030

Waid-Ebbs, J. K., Wen, P.-S., Heaton, S. C., Donovan, N. J., & Velozo, C. (2012, Dec 2012). The item level psychometrics of the behaviour rating inventory of executive function-adult (brief-a) in a TBI sample. *Brain injury, 26*(13-14), 1646-1657.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale--Fourth Edition (WAIS-IV)*. APA PsycTests.

Wechsler, D. (2009). *WMS-IV: Wechsler memory scale*. Pearson.

Wefel, J. S., Parsons, M. W., Gondi, V., & Brown, P. D. (2018). Neurocognitive aspects of brain metastasis. *Handb Clin Neurol*, *149*, 155-165. https://doi.org/10.1016/b978-0-12-811161-1.00012-8

World Health Organization, W. (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020*

CHAPTER III: PUBLIC DISSEMINATION DOCUMENT FOR THE META-ANALYTIC REVIEW

Background

A traumatic brain injury (TBI) is an alteration of brain function or pathology caused by an external force (Menon et al., 2010) and are often associated with executive dysfunction (Stuss et al., 1985). Executive functions represent an important domain of abilities that are important for purposeful goal-directed behaviour (such as attention, planning and organisation and initiation of activity) (Donders et al., 2015). Previous findings have suggested a link between executive functioning and cognitive recovery, although the evidence for this is mixed. Others have suggested a link between executive functioning and functional outcomes (Allanson et al., 2017), driving outcomes and employment outcomes following a TBI.

What did the study do?

A total of 720 articles were found following a search from the EMBASE, PsychInfo and MEDLINE databases. A further two articles were hand-searched from references. Twenty-four met inclusion criteria and were included in this meta-analytic review. The criteria for inclusion was that interventional studies could be included if they recorded relevant outcome data, participants included were drawn from participants with a TBI, outcome data required include means and standard deviations, F-test, Cohen's d or an r effect size reported. Studies also had to report using one of the following executive function measures; the Wisconsin Card Sorting Test, the Trail Making Test, or the Verbal Fluency Test and one of the following outcome measures; the Glasgow Outcome Scale, the Disability Rating Scale, the Mayo Portland Adaptability Inventory or the Community Integration Questionnaire. Included studies also exploring employment outcomes and driving outcomes. Studies were excluded if they were a meta-analysis, review, commentary or conference abstract. Also, studies were excluded if the sample size was less than ten.

What did the meta-analysis find?

This meta-analytic review found that tests of executive functioning, specifically the Trail Making Test and Wisconsin Card Sorting Test, were significantly associated with functional outcomes following a TBI. Verbal Fluency was not significantly associated with functional outcome following a TBI. However, there was a high level of heterogeneity in the studies included in the analysis suggesting that the estimates of associations between test of executive functioning and functional outcome may be biased to the presence of confounding between studies factors. Further analysis focused on identifying the source of heterogeneity.

The only test of executive functioning that was entered into this analysis (due to a scarcity of research in this domain) was the Trail Making Test. This analysis concluded that the Trail Making Test was also associated with a person's ability to return to driving following a TBI, although there were only three studies reporting on this outcome, thus limiting any conclusions drawn from this analysis.

Regarding returning to employment following a TBI, this meta-analytic review found that no test of executive functioning was associated with employment outcomes following a TBI. However, this analysis only included outcomes from four studies, therefore limiting any conclusions drawn from this analysis.

What do the results mean?

This meta-analytic review found that tests of executive functioning, specifically the Trail Making Test and the Wisconsin Card Sorting Test, were associated with functional outcomes following a TBI, therefore confirming the previous meta-analytic review by Allanson et al. (2017). This has important implications for clinicians, especially as the findings can help to guide rehabilitation strategies. Moreover, these findings may be used to help individuals plan

for future care needs. This meta-analytic review found that the Trail making Test was associated with predicting whether an individual will return to driving following a TBI, however no association was found between test of executive functioning and employment outcomes. Importantly, this meta-analytic review has highlighted the scarcity of research in specific outcomes (such as employment and driving), which offers an avenue for future research.

References for the Public dissemination document for the meta-analytic review

Allanson, F., Pestell, C., Gignac, G. E., Yeo, Y. X., & Weinborn, M. (2017). Neuropsychological predictors of outcome following traumatic brain injury in adults: a meta-analysis. *Neuropsychology Review, 27*(3), 187-201.

Donders, J., Oh, Y. I., & J., G. (2015). Self- and Informant Ratings of Executive Functioning After Mild Traumatic Brain Injury. *J Head Trauma Rehabil . 30*(6), E30-E39.

Menon, D. K., Schwab, K., Wright, D. W., & Maas, A. I. (2010). Demographics and clinical assessment working group of the International and interagency initiative toward common data elements for research on traumatic brain injury and psychological health. Position statement: definition of traumatic brain injury. *Arch Phys Med Rehabil, 91*(11), 1637-1640.

Stuss, D., Ely, P., Hugenholtz, H., Richard, M., LaRochelle, S., Poirier, C., & Bell, I. (1985). Subtle neuropsychological deficits in patients with good recovery after closed head injury. *Neurosurgery, 17*(1), 41-47.

CHAPTER IV: PUBLIC DISSEMINATION DOCUMENT FOR THE EMPIRICAL PAPER

Background

The growth in telemedicine was enhanced in March 2020, when the World Health Organisation announced the outbreak of the COVID-19 pandemic to the world (World Health Organization, 2020). As a result, many clinics were suspended or moved to virtual consultations and assessment. There was an increase in reliance of videoconferencing software to continue providing essential patient care, and this was supported by the British Psychological Society (BPS) Division of Neuropsychology (DON) and included the provision of care from neuropsychologists. Given the reliance of test batteries in the neuropsychological assessment, it is essential to understand whether these tests, when administered virtually, provide a valid and equivalent assessment of an individual's cognitive functioning, to if they had received the assessment face-to-face. Previous research has yielded mixed findings, with some finding a virtual neuropsychological assessment to yield valid results (Gnassounou et al., 2021), whereas others have found differences in performance when administering the assessment virtually compared with a face-to-face assessment (Brearly et al., 2017).

What did the study do?

28 healthy participants took part in the study. A counterbalancing design was employed, whereby each participant completed the neuropsychological assessment twice, once virtually and once face-to-face with half completing the virtual assessment first, and the other half completing the face-to-face assessment first. The neuropsychological assessment battery assessed an individual's general intellectual functioning, memory and attention, executive functioning (encompassing domains such as orientation, planning and inhibition), information processing speed and language functioning. An assessment of their effort was also administered as a standalone measure, as well as embedded within the test material.

What did the study find?

This study found that there was no significant difference between face-to-face administration of many of the neuropsychological assessments that were used to make up this test battery compared with virtual administration. However, there were significant differences in mode of administration for some of the executive functioning tasks. There were significant differences on the DKEFS Colour Naming task (Delis et al., 2001), with participants making fewer errors on the colour naming task and inhibition/switching task when administered virtually compared to face-to-face administration. There was also a significant age cohort effect in the inhibition/switching task (Delis et al., 2001).

What do the results mean?

The administration of a battery of neuropsychological assessments virtually largely provide a valid alternative to face-to-face assessments. However, some of the assessments, particularly the tests of executive functioning, may not produce a valid assessment of an individual's executive functioning when administered virtually compared with face-to-face administration, therefore caution must be exercised when interpreting these tests. Consideration must also be given to test selection, as well as practical considerations (such as internet stability) and screen size. Moreover, the present findings may not translate to a clinical population, therefore further research is required to replicate these findings in a clinical setting with specific patient groups. Other important considerations must focus on preserving the security and integrity of test materials, especially in a medico-legal setting as well as reconciling current administration guidelines of some neuropsychological assessments (such as a technician must be present in the room during the assessment) with virtual administration of test material. Future research should also focus on developing a neuropsychological assessment battery using information technology, which may have implications for not only providing a more accurate and sensitive

recording of an individual's test performance, but also save time in scoring and analysing an individual's assessment.

References for the press release document for the empirical study

Brearly, T. W., Shura, R. D., Martindale, S. L., Lazowski, R. A., Luxton, D. D., Shenal, B. V., & Rowland, J. A. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology Review, 27*(2), 174-186.

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

Gnassounou, R., Defontaines, B., Denolle, S., Brun, S., Germain, R., Schwartz, D., Schück, S., Michon, A., Belin, C., & Maillet, D. (2021). Comparison of Neuropsychological Assessment by Videoconference and Face to Face. *Journal of the International Neuropsychological Society*, 1-11. https://doi.org/10.1017/S1355617721000679

World Health Organization, W. (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020*

Appendix 1: Poster to advertise the project

UNIVERSITY OF BIRMINGHAM

Evaluation of convergent validity of face-to-face and teleconference neuropsychological assessment

Are you **aged between 18 and 35**?

Are you a **Native English Speaker** without a diagnosis of a neurological condition or Learning Disability?

Are you interested in taking part in research?

What is the study investigating?

Despite the increase in telemedicine, there have been few studies that have assessed the validity of administering these tests using videoconferencing software. This study aims to assess the convergent validity of face-to-face and teleconference neuropsychological assessments.

What will the study involve?

- It will take part at the University of Birmingham
- You will be asked to complete a battery of neuropsychological assessments both face-to-face and remotely using a method of tele-communication (such as Skype, Zoom or MS Teams)

How do I get involved?

If you are interested in taking part, please contact:

Dr Carl Krynicki (Trainee Clinical Psychologist)

[redacted]

School of Psychology,
University of Birmingham

Version 1
23/09/2021

Appendix 2: Participant information sheet

UNIVERSITY OF BIRMINGHAM

Evaluation of convergent validity of face-to-face and teleconference neuropsychological assessment

Participant Information Sheet

We would like to invite you to take part in a research study. Before making any decisions, you need to understand why this research is being done and what the study would involve for you. Please read the following information carefully and, once you have finished, feel free to ask any questions. Take time to decide whether or not you wish to take part.

What is the purpose of the study?

Neuropsychological tests assess cognitive functioning (including planning, memory, mental flexibility and inhibition). The use of video-conferencing to administer neuropsychological assessments has received much attention over the last 10 years, but this has steeply accelerated over the last year. Despite the increase in telemedicine, there have been few studies that have explored whether the results for assessments administered using tele-communication software yield valid and reliable results. Therefore, the aim of this study is to assess the validity of neuropsychological assessments administered using video-conferencing software.

Why have I been chosen?

You have been invited to take part in this study because you are a native English speaker, between the ages of 16 and 89 without a diagnosis of a neurodegenerative condition (such as dementia) or learning disability. To be eligible to take part in the study, you must be able to give informed consent.

What will happen to me if I agree to take part?

If you decide to take part, you will be asked to provide some demographical information and then to complete a battery of neuropsychological assessments with a trained researcher. The neuropsychological assessments will test your cognitive abilities including your attention, concentration, memory, planning, organisation and visual processing. The assessment should take around an hour and a half to complete. You will then be asked to complete the assessments again at least three days later, using video-conferencing software (Zoom, Skype or Microsoft Teams) or in person depending on how the first assessment was administered.

Do I have to take part?

No, you do not have to take part – it is completely up to you to decide. If you agree to participate, however, we will ask you to sign a consent form to show that you have read and understood the information sheet. You can withdraw your participation at any point throughout the study. If you decide to withdraw, you do not have to offer an explanation and your decision will, in no way, affect your future treatment. If you decide to withdraw, you can withdraw your data until data analysis has begun, which will be the 21st of March 2022. It may not be possible to withdraw data once the analysis has begun.

What are the benefits of taking part?

UNIVERSITY OF BIRMINGHAM

There may be no direct benefit to you as a consequence of participating in the study; however, the results may help improve the treatment of other people in the future.

What are the possible disadvantages to taking part?

There are no immediate disadvantages to taking part.

The University has in force a Public Liability Policy and/or Clinical Trials policy which provides cover for claims for "negligent harm" and the activities here are included within that coverage

What will be done to ensure confidentiality?

All data will be anonymised, and personal details that might identify you will be removed so that you cannot be identified in any published reports. Data will be stored on a secure database and hardcopies kept in a locked cabinet for 10 years (the minimum time stipulated by the University of Birmingham). Only members of the research team will have access to this data and relevant others at the University of Birmingham to ensure that the analysis is a fair and reasonable representation of the data.

What if there is a problem?

If you have found some of the topics discussed in this research difficult, then we encourage you to contact your GP or Samaritans (116 123) for further information and support.

If you have any questions regarding the study, you can contact Dr Carl Krynicki on [redacted]

If you are unhappy or wish to make a complaint, please contact Mrs Sue Cottam, Research Ethics Manager, [redacted] All concerns will be dealt with promptly, and information will subsequently be provided either by telephone or in writing to inform you of how the problem has been addressed.

What will happen to the results of the study?

The data obtained from this study will be analysed and submitted as part of Dr Carl Krynicki's Clinical Psychology doctorate degree. We also hope to publish the results in scientific journals. If you would like to, you will be provided with a copy of the final published article.

How will we use information about you?

We will need to use information from you for this research project.

This information will include your initials, name, contact details, responses to the questionnaires and discussion with your responsible clinician. People will use this information to do the research or to check your records to make sure that the research is being done properly.

People who do not need to know who you are will not be able to see your name or contact details. Your data will have a code number instead.

We will keep all information about you safe and secure.

UNIVERSITY OF BIRMINGHAM

Once we have finished the study, we will keep some of the data so we can check the results. We will write our reports in a way that no-one can work out that you took part in the study.

What are your choices about how your information is used?

You can stop being part of the study at any time, without giving a reason, but we will keep information about you that we already have.

We need to manage your records in specific ways for the research to be reliable. This means that we will not be able to let you see or change the data we hold about you.

Where can you find out more about how your information is used?

You can find out more about how we use your information;

- at www.hra.nhs.uk/information-about-patients/
- our leaflet available from **www.hra.nhs.uk/patientdataandresearch**
- by asking one of the research team
- by sending an email to dataprotection@contacts.bham.ac.uk

Who has reviewed this study?

This study has been reviewed by a Research Ethics Committee (ERN_21-1412).

Thank you for taking the time to read this information sheet. If you would like further information or have any questions, please contact:

Dr Carl Krynicki
[redacted]
School of Psychology,
University of Birmingham,
52 Pritchatts Road,
Birmingham
B15 2SA

Appendix 3: Consent form

UNIVERSITY OF BIRMINGHAM

Evaluation of convergent validity of face-to-face and teleconference neuropsychological assessment

Participant Consent Form

Chief Investigator: Dr Carl Krynicki, Principal Investigator: Dr Christopher Jones, Dr David Hacker

Participant Number ……

Consent for participation in the above study:

1. I can confirm that I have read and understood the participant information sheet (version 3.1), dated 09/11/2021, for the study and have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	Initials …….
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without reason or penalty, without my legal rights being affected.	Initials …….
3. I understand that data collected during the study may be looked at by other members of the research team and relevant others at the University of Birmingham to ensure that the analysis is a fair and reasonable representation of the data.	Initials …….
4. I understand that the data will be stored anonymously (during and after the study). This means that the data will not be able to be traced back to me.	Initials …….
5. I agree that the information collected from me can be examined and stored for up to 10 years at the study sites. This is in accordance with the University of Birmingham's data handling guidelines.	Initials …….
6. I agree to take part in the above study.	Initials …….
7. I agree to not record any part of the assessment.	Initials …….
8. I would like to be informed of the results. Declining to do so will not affect my participation in the study in any way.	**YES/NO** Initials …….

Name of participant …… Signed …… Date ……

Name of researcher …… Signed …… Date ……

Two wet ink consent forms; one to retain in the site file, and one for the participant

Consent form
Version 3.1
09/11/2020
IRAS no: 305388

Appendix 4: Test of Memory Malingering

Test material removed due to copyright. For more information about this, please visit: https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Test-of-Memory-Malingering/p/100000191.html#:~:text=Based%20on%20research%20in%20neuropsychology,malingering%20from%20genuine%20memory%20impairments.

Appendix 5: WAIS-IV Digit Span

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Adult-Intelligence-Scale---Fourth-UK-Edition/p/P100009273.html?gclid=CjwKCAjw682TBhATEiwA9crl31cXu3YPRTEITTDrew78VMTGbB6bHwWHPCUO2y3Y7R6_YRAxSSMxyBoC0RAQAvD_BwE

Appendix 6: Embedded effort tests from the WMS-IV Logical Memory and Visual Reproduction

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Adult-Intelligence-Scale---Fourth-UK-Edition/p/P100009273.html?gclid=CjwKCAjw682TBhATEiwA9crl31cXu3YPRTEITTDrew78VMTGbB6bHwWHPCUO2y3Y7R6_YRAxSSMxyBoC0RAQAvD_BwE

https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Memory-Scale---Fourth-UK-Edition/p/P100009265.html?gclid=CjwKCAjw682TBhATEiwA9crl3zgJzOyUJvedYid5NXIeNVve22fXPsyHjG-5ZAvuQSaWTpKQvD9B_xoCZ-IQAvD_BwE

Appendix 7: WAIS-IV Test of Premorbid Functioning

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Adult-Intelligence-Scale---Fourth-UK-Edition/p/P100009273.html?gclid=CjwKCAjw682TBhATEiwA9crl31cXu3YPRTEITTDrew78VMTGbB6bHwWHPCUO2y3Y7R6_YRAxSSMxyBoC0RAQAvD_BwE

Appendix 8: WAIS-IV Information

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Adult-Intelligence-Scale---Fourth-UK-Edition/p/P100009273.html?gclid=CjwKCAjw682TBhATEiwA9crl31cXu3YPRTEITTDrew78VMTGbB6bHwWHPCUO2y3Y7R6_YRAxSSMxyBoC0RAQAvD_BwE

Appendix 9: WAIS-IV Matrix reasoning

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Adult-Intelligence-Scale---Fourth-UK-Edition/p/P100009273.html?gclid=CjwKCAjw682TBhATEiwA9crl31cXu3YPRTEITTDrew78VMTGbB6bHwWHPCUO2y3Y7R6_YRAxSSMxyBoC0RAQAvD_BwE

Appendix 10: WMS-IV Logical Memory

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Memory-Scale---Fourth-UK-Edition/p/P100009265.html?gclid=CjwKCAjw682TBhATEiwA9crl3zgJzOyUJvedYid5NXIeNVve22fXPsyHjG-5ZAvuQSaWTpKQvD9B_xoCZ-IQAvD_BwE

Appendix 11: WMS-IV Visual Reproduction

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Memory-Scale---Fourth-UK-Edition/p/P100009265.html?gclid=CjwKCAjw682TBhATEiwA9crl3zgJzOyUJvedYid5NXIeNVve22fXPsyHjG-5ZAvuQSaWTpKQvD9B_xoCZ-IQAvD_BwE

Appendix 12: Rey-Osterreith Complex Figure

Test material removed due to copyright. For more information, Please visit: https://www.parinc.com/products/pkey/127

Appendix 13: Oral Trail Making Test

Test material removed due to copyright.

Appendix 14: Verbal Fluency Test

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/delis/Delis-Kaplan-Executive-Function-System/p/P100009078.html

Appendix 15: Colour-Word Interference Test

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/delis/Delis-Kaplan-Executive-Function-System/p/P100009078.html

Appendix 16: Hayling Sentence Completion Test

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Executive-Function/Hayling-and-Brixton-Tests/p/P100009219.html

Appendix 17: Brixton Spatial Anticipation Test

Test material removed due to copyright. For more information, please visit: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Executive-Function/Hayling-and-Brixton-Tests/p/P100009219.html

Appendix 18: Oral symbol digit modalities test

Test material removed due to copyright. For more information, please visit: https://www.annarbor.co.uk/index.php?main_page=index&cPath=416_249_306

Appendix 19: Boston Naming Test

Test material removed due to copyright. For more information, please visit: https://www.annarbor.co.uk/index.php?main_page=product_info&products_id=1686

Appendix 20: Ethical approval

Application for Ethical Review ERN_21-1412

SW Samantha Waldron (Research Support Services)
To Christopher Jones (Psychology)
Cc Carl Krynicki (ClinPsyD Clinical Psychol FT); 'David Hacker'

Reply | Reply All | Forward

Tue 09/11/2021 13:29

Action Items | Get more add-ins

Dear Dr Jones,

Re: "Evaluation of convergent validity of face-to-face and teleconference neuropsychological assessment"
Application for Ethical Review ERN_21-1412

Thank you for your application for ethical review for the above project, which was reviewed by the Science, Technology, Engineering and Mathematics Ethical Review Committee.

On behalf of the Committee, I confirm that this study now has full ethical approval.

I would like to remind you that any substantive changes to the nature of the study as described in the Application for Ethical Review, and/or any adverse events occurring during the study should be promptly brought to the Committee's attention by the Principal Investigator and may necessitate further ethical review.

Please also ensure that the relevant requirements within the University's Code of Practice for Research and the information and guidance provided on the University's ethics webpages (available at https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Links-and-Resources.aspx) are adhered to and referred to in any future applications for ethical review. It is now a requirement on the revised application form (https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Ethical-Review-Forms.aspx) to confirm that this guidance has been consulted and is understood, and that it has been taken into account when completing your application for ethical review.

Please be aware that whilst Health and Safety (H&S) issues may be considered during the ethical review process, you are still required to follow the University's guidance on H&S and to ensure that H&S risk assessments have been carried out as appropriate. For further information about this, please contact your School H&S representative or the University's H&S Unit at healthandsafety@contacts.bham.ac.uk.

Kind regards,

Ms Sam Waldron (she/her)
Research Ethics Officer
Research Support Group
University of Birmingham
Email: [redacted]
Video/phone: If you would like to arrange a Teams/Zoom/telephone call, please email me and I will get in touch with you as soon as possible.
Postal address: Ms Sam Waldron, Finance Office, University of Birmingham, c\o Room 106 Aston Webb, B Block, Edgbaston, Birmingham, B15 2TT.

Web: https://intranet.birmingham.ac.uk/finance/RSS/Research-Support-Group/Research-Ethics/index.aspx
Click Research Governance for further details regarding the University's Research Governance and Clinical Trials Insurance processes, or email researchgovernance@contacts.bham.ac.uk with any queries relating to research governance.

Appendix 21: Approval of risk assessment

SOPHS_21_100_CJ Evaluation of convergent validity of face-to-face and teleconference neuropsychological assessment

Reply | Reply All | Forward

DC Denise Clissett (Psychology)
To Christopher Jones (Psychology); Carl Krynicki (Psychology)
Cc Massimiliano Di Luca (Psychology)
Thu 09/12/2021 11:51

Action Items | Get more add-ins

Hi Chris and Carl
I can confirm that your RA SOPHS_21_100_CJ for your study Evaluation of convergent validity of face-to-face and teleconference neuropsychological assessment has been approved under general RA SOPHS_20_01_JC.

Can I remind you of the 24hr before pre screening need to be adhered to as we need to know that our participants are well before they come into the building.

PPE can be collected from me, can you be in touch to collect it in advanced as I'm not in every day

Any further correspondence about this RA number; please use this thread or quote the RA SOPHS_21_100_CJ
Thank you

Regards
Denise

Denise Clissett
School of Psychology
College of Life & Enviromental Sciences
University of Birmingham

PA Support to Professor Matthew Broome
Institute for Mental Health | University of Birmingham

Find us at: @IMH_UoB | www.birmingham.ac.uk/imh

UNIVERSITY OF BIRMINGHAM | imh institute for mental health

School Health & Safety Coordinator.
School Patient Coordinator

[redacted]

Working remotely from home Tue/Wed/Thur
Office based Mon/Fri.