The validity of neuropsychological assessment tools for the assessment of general and domain-specific cognitive functioning after an acquired brain injury

Thesis submitted in part fulfilment of the degree of

Doctorate in Clinical Psychology (DClinPsy)

University of Leicester

By

Corran Moore

Supervised by Dr Gerald Burgess and Dr Alice Welham

August 2020

Declaration

I confirm that the literature review and empirical research reported in this thesis are original pieces of work. They are submitted in part fulfilment of the Doctorate in Clinical Psychology (DClinPsy) and have not been submitted for any other academic award. This thesis was checked for completion prior to submission.

The validity of neuropsychological assessment tools for the assessment of general and domain-specific cognitive functioning after an acquired brain injury

By Corran Moore

Thesis Abstract

Accurate assessment and categorisation of cognitive impairment after an acquired brain injury is vital for the creation of targeted rehabilitation plans and the planning and provision of care. To be useful neuropsychological assessment tools must be able to predict patient outcomes and accurately classify the presence/absence of impairment.

Literature Review

The ability of domain-specific assessment tools used post-stroke, to predict later functional outcomes was reviewed. Tools assessing the domains of executive function, visuospatial perception/construction and visual memory were selected for review, due to previous evidence of their relationship with functional outcomes. It was concluded that all domains were predictive of a range of functional outcomes at many time-points post-stroke. Individual assessment tools showing evidence of predictive validity were identified for clinical use.

Empirical Report

This study assessed the diagnostic ability of the Short Parallel Assessments of Neuropsychological Status (SPANS) in acquired brain injury (ABI) samples. The ability of the SPANS to accurately distinguish between a sample of patients who had suffered an ABI and a sample of age and education matched healthy norms was investigated using five different clinical samples: ≤ 6-months post-ABI, ≥ 1-year post-ABI, high school educated, college/vocationally educated and university educated. The SPANS total score and individual indices had high diagnostic validity in the acute ABI sample, indicating it is a valid general and domain-specific screen in this setting. The /SPANS total score also had high diagnostic validity in the long-term ABI sample indicating it is a valid general cognitive screen in this setting. Scores for individual indices in this sample were mixed, this could be because of recovery of functioning in particular domains or a lack of sensitivity in subtests. Education level and time-since injury had a large impact on optimal cut-off scores, the implications for the use of 'one-size-fits-all' cut-offs is discussed.

Acknowledgements

I would like to thank my supervisor, Dr Gerald Burgess, for all of his expertise, support and encouragement. His advice and judgement throughout the conception and development of this research have been invaluable and this project would not have been possible without him.

I would also like to thank my second supervisor, Dr Alice Welham, for taking the time out of her busy schedule to read through my drafts and to offer her very valuable advice and insights. Thank you also for all of your encouragement and for always being there at the end of the phone.

Lastly, I would like to thank my friends and family for supporting me, keeping me calm and always believing in me – even when I didn't. A special thank you to my partner, Sean, for his unwavering support and many cups of tea!

Word Counts

Thesis Abstract:	**299**
Part One: Systematic Literature Review	
Abstract	**225**
Full text (excluding figures and tabulated data)	**7196**
References	**1139**
Part Two: Empirical Report	
Abstract	**417**
Full text (excluding figures and tabulated data)	**7202**
References	**708**
Non-mandatory Appendices:	**3524**
Total word count for thesis	**18, 564**
(excluding mandatory appendices and references)	

Contents **Page**

List of Appendices

List of Tables

Part One: Systematic Literature Review

Part Two: Empirical Report

List of Figures

Part One: Systematic Literature Review

Part Two: Empirical Report

Part One: Systematic Literature Review[1]

A review of the predictive relationship between domain-specific assessment in the domains of visual memory, visuospatial perception/construction, visuomotor speed and executive function post-stroke and later functional outcomes

[1] This review has been formatted for submission to the Journal of Neuropsychological Rehabilitation, journal guidelines are listed in Appendix A.

Abstract

Objectives: To establish whether domain-specific assessment of executive function (EF), visual memory (VM) visuospatial perception/construction (VSPC) and visuomotor speed (VMS), post-stroke, is predictive of different functional outcomes, and whether this relationship persists across multiple assessment time-points. An additional aim was to identify domain-specific assessment tools showing predictive validity for functional outcomes post-stroke.

Method: A systematic search of databases was conducted to identify relevant papers. Papers were then screened for duplicates and compared to inclusion/exclusion criteria. A total of 17 papers were identified for review.

Results: Domain-specific testing in the domains of VSPC and EF were found to be predictive of multiple functional outcomes. The domains of VM and VMS were both deduced to be predictive of later functioning in Activities of Daily Living. These relationships were inferred for multiple assessment timepoints. Twelve domain-specific assessment tools were concluded to show predictive validity of functional outcomes post-stroke.

Conclusions: Functioning in the cognitive domains of EF, VSPC, VM and VMS post-stroke, in both acute and later stages, is predictive of long-term functional outcomes. More research is needed to see if these relationships differ based on cognitive domain and specific outcome. There was evidence that the particular domain-specific assessment tool used mattered, with some being predictive while other were not. Clinical assessment tools should therefore be chosen based on individual evidence of validity rather than the domain they assess.

1. Introduction

1.1 Stroke and cognitive impairment

Stroke is one of the leading causes of disability and loss of independence worldwide (Seshadri & Wolf, 2007), with patients who have experienced a stroke spending more days in hospitals and care homes than people with any other medical condition (Scott & Scott, 1994). Research has found that generalised cognitive impairment post-stroke is one of the most important predictors of later functional recovery (Ballard et al., 2003; Barker-Collo & Feigin, 2006).

Cognitive impairment post-stroke, however, is rarely generalised. A typical stroke presentation is a pattern of damaged and preserved functioning, which can differ dramatically between individuals, depending on the type of stroke and location of damage (e.g. Teasell & Norhayati, 2016). Researchers have argued that impairments in different cognitive domains are unlikely to have the same impact on functional outcomes, with some likely to be more disabling overall and some more disabling for specific areas of functioning than others (Middleton et al., 2014; Nys et al., 2005). Despite this, the relationship between domain-specific cognitive impairments and functional outcomes has been far less studied and is subsequently not so well established.

Establishing which areas of cognitive impairment post-stroke are related to specific functional outcomes is important as it helps to inform the potential utility of domain-specific cognitive assessment tools and provide information needed to aid the development of targeted, evidence-based interventions and inform clinical care decisions, including the prioritisation and commissioning of services and support.

1.2 Previous Reviews

A previous review by Barker-Collo and Feigin (2006) aimed to establish the impact of different types of neurological and cognitive impairment on functional outcomes post-stroke. They found that research was sparse for many cognitive areas, with many studies focusing on neuropsychological disorders such as neglect or only on single cognitive domains, most commonly attention or executive function. Due to insufficient research into different areas of cognitive impairment in the domain of stroke they had to include data from geriatric and traumatic brain injury populations. Several methodological issues were also found to be common, with much of the research being cross-sectional or short-term and therefore unable to establish long-term effects. Additionally, sample sizes were often small and non-population-based, meaning generalisability was potentially low and the possibility of bias high.

They concluded that neglect, anosognosias and aphasias showed promise for predicting functional outcomes, as well as deficits in the cognitive domains of verbal memory and attention (Barker-Collo & Feigin, 2006). However, they felt that there was insufficient data to establish links between other cognitive domains and functional outcomes and that more rigorous, population-based research was needed to establish these relationships. A major flaw of this review was that studies were only included if they separated results based on lesion side, due to this being another area of interest for the review paper. This decision has debatable theoretical grounding and was one of the main contributing factors to the limited research available, limiting the scope and generalisability of the review.

Fourteen years on, the relationship between neglect and later functional outcomes has been well established and reviewed (e.g. Di Monaco et al., 2011; Nijboer et al., 2013). However, the relationship between domain-specific cognitive impairments and long-term

functional outcomes is still unclear. A recent review by Mole and Demeyere (2020) aimed to clarify the relationship between both domain-general and domain-specific assessments post-stroke and later functional outcomes. Outcomes were categorised as 'activities' ("learning and applying knowledge," "general tasks and demands," "communication," "mobility," and "self-care") and 'participation' ("domestic life," "interpersonal interactions and relationships," "major life areas," and "community, social and civic life") with outcome measures coded and categorised based on which of these categories the majority of their items fell under. They found that domain-specific tests were more reliable predictors of later outcomes than domain-general assessments and concluded that testing in the cognitive domains of visuospatial perception/construction, visual memory, visual neglect, and attention/executive function showed predictive validity for functional outcomes at 6–12 months post-stroke, specifically for 'activities'. Insufficient evidence was available to draw similar conclusions for 'participation'.

This review was methodologically sound, using appropriate quality appraisal tools and well-justified inclusion/exclusion criteria. One of these criteria was strict assessment time-points, with cognitive assessments needing to be completed within the first 6 weeks of injury and outcomes assessed between 6-12 months post injury. This was based on clinical guidance that cognitive testing should be undertaken in the acute phase post-injury, and the recommendation that health needs are assessed at 6-12 months post-stroke. Whilst these strict inclusion criteria are useful for increasing the potential validity of conclusions, in relation to specific circumstances compatible with clinical guidelines, they limit the generalisability of conclusions to different clinical settings. In clinical practice cognitive assessments often take place several months, or even years post-injury and cognitive impairment often endures long-term (Blomgren et al., 2019). Information regarding the predictive value of domain-specific functioning at later time-points is therefore clinically relevant.

An additional impact of these strict inclusion criteria was a limitation on the number of studies meeting criteria for review, with between one and three studies identified for each domain, and only one of these finding a significant relationship in some cases. This could impact on the validity and generalisability of conclusions drawn and is likely also responsible for the slightly limited scope in terms of specific functional outcomes. Outcomes were categorised into two broad areas in order for a relatively small number of studies, with diverse outcome measures, to be amalgamated. Clinically relevant differences in the relationship between cognitive domains and specific outcomes may have been lost, for example, we cannot know if visuospatial impairments were specifically predictive of difficulties performing activities of daily living or problems with social relationships. Specific functional outcomes such as employment were also not investigated. These different relationships are vital to the utility of information provided by cognitive assessments for rehabilitation planning. If the aim is to help a patient back into work for example, then knowing which cognitive areas are most predictive of employment outcomes would be necessary for the selection of appropriate assessment tools. This is particularly important given that evidence, already stated above, indicates relationships between particular cognitive domains and specific functional outcomes are likely different (e.g. Nys et al., 2005).

An additional area to consider is evidence of the validity of specific assessment tools, such as particular batteries, subtests and composites, rather than just domains of cognition. The distinction between cognitive assessment tools and cognitive domains is an important one as it is virtually impossible for assessments to measure one domain to the exclusion of all others. For example, tests of visuospatial ability often also involve processing speed and psychomotor functioning. Analysis of the relationship between scores on these tests and later outcomes would not be able to distinguish which combination of these processes were responsible for the predictive value of the test.

Cognitive domains are also broad, with two tests of 'executive functioning' potentially measuring quite different processes, e.g. inhibition vs initiation, which may be differentially predictive. Using results from one assessment tool, or composite of tools (often with no internal validity statistics reported) to form conclusions around the predictive value of a whole cognitive domain can therefore be misleading; just because one test of executive function is linked to functional outcomes it does not follow that all assessment tools in this domain will be. The importance of avoiding over-reliance on the categorisation of 'domain-specific' tests is further demonstrated by the fact that the same cognitive subtests are used in research as measures of different cognitive domains. The Trail-Making Test (TMT A & B; Reitan, 1955), for example, has been separately categorised and used, in different research papers, as a test of visual attention, executive function and visuomotor ability; in reality it assesses all of these skills (Sánchez-Cubillo et al., 2009). Conclusions based on specific tests or composites may therefore have higher clinical validity allowing clinicians to select cognitive assessment tools with proven predictive validity.

1.3 Aims

The current review aims to address these areas by focusing on the cognitive domains found to be predictive of 'activities' in the previous review, executive function/attention, visuospatial ability and visual memory, but extending inclusion/exclusion criteria to include a wider range of assessment timepoints and functional outcome measures. The main aims of the review are:

1) To summarise the relationship between domain-specific assessment in the cognitive domains of visuospatial perception/construction, visual memory, and attention/executive functioning, for specific functional outcomes (e.g. employment,

activities of daily living, quality of life and social relationships). This will be examined at a range of timepoints to indicate whether conclusions are relevant to cognitive assessments performed at both early and later time-points and long-term functional outcomes.

2) To summarise evidence for the predictive validity of individual domain-specific assessment tools for later functional outcomes post-stroke.

2 Method

2.1 Search strategy

A systematic search of Web of Science, Medline, PsycINFO and Scopus was conducted using the following search terms, developed in an initial scoping search:

1) Stroke OR "Cerebro* accident" OR CVA OR "Cerebro* Event" OR CVE OR "Cerebro* Incident" OR CVI

AND

2) Cognit* OR Neuropsychologic* OR "Executive function*" OR Visuo* OR Attention OR Memory OR Dysexecutive OR "Processing Speed" OR "Psychomotor Speed" OR Percep*

AND

3) Outcome OR Functional OR Functioning OR "Quality of Life" OR QoL OR Rehabilitat* OR "Activities of Daily Living" OR ADL OR Employ*

These databases were chosen in order to capture a comprehensive selection of research from different disciplines, including medicine, social science, health science and psychology, all of which are relevant to the subject area being reviewed.

Search terms were kept deliberately broad and searching was done at an abstract level, as initial scoping had revealed that specific cognitive domains and functional outcomes were often used alongside a multitude of other independent/dependent variables and were not always included in titles or abstracts.

2.2 Selection criteria

The following inclusion criteria were applied:

- Original, peer-reviewed article published in the English language
- Human participants over the age of 18 who had suffered a stroke or TIA
- Longitudinal research
- Including results assessing the relationship between testing in the domains of executive function, visual memory, or visuospatial perception/construction, undertaken at a baseline time-point, and functional outcomes at a later time-point.
- Articles were excluded if they were reviews, meta-analyses or unpublished theses.

2.2.1 Inclusion criteria for assessment time-points

Inclusion criteria for testing time-points were kept broad in order to increase the generalisability of conclusions. Any form of longitudinal research were included. Studies where cognitive testing was undertaken in the acute inpatient phase, and outcomes were assessed at discharge were included due to the steep improvement in cognitive function usually found over this period (Wade, Wood, & Hewer, 1985). All papers where cognitive assessment was completed in the post-acute phase had at least 6 months between cognitive testing and assessment of functional outcomes Inclusion criteria for cognitive assessments

Papers were included only if the cognitive assessments used were standardised neuropsychological assessments, subtests or composites published or available in the public

domain. These exclusion criteria were to ensure validity and generalisability of results, and to allow conclusions to be drawn about the predictive validity of specific subtests and composites that are oft used in clinical services. This resulted in some articles which were included in the previous review being screened out. Articles were included if tests not meeting the inclusion criteria were used in combination with ones that did, and these data could be extracted due to reporting of test-specific statistics. For inclusion in the review cognitive assessment needed to take place in the cognitive domains of visuospatial perception/construction, executive function or visual memory. Executive function (EF)

Including the following areas of processing:

- Task-switching e.g. the Trail-Making Test B (TMT-B)
- Planning/strategising e.g. the Wisconsin Card Sorting Task (WCST)
- Behavioural regulation e.g. the Behavioural Dyscontrol Scale (BDS)
- Initiation e.g. the Verbal Fluency Test (VFT)
- Response inhibition e.g. the Stroop Test

Tests of attentional switching such as the Visual Elevator Test and TMT-A also met inclusion criteria for this domain as there is significant overlap between these tests and the domain of EF. Tests of other types of attention, such as selective or general, were not included.

2.2.1.1 Visual Memory (VM)

Including the following areas:

- Picture recall e.g. Weschler Memory Scale Picture Recall Subtests
- Delayed visual reproduction e.g. the Rey-Osterrieth Complex Figure-delayed (ROCF-delayed)
- Location learning e.g. the Location Learning Task

2.2.1.2 Visuospatial Perception/Construction (VSPC)

Including the following areas:

- Judgement of orientation e.g. the RBANS Judgement of Line Orientation Subtest
- Visual picture copy e.g. the Rey-Osterrieth Complex figure-copy (ROCF-copy)
- Shape construction e.g. the WAIS Block Design (WAIS BD)
- Facial recognition e.g. the Test of Facial recognition

Tests of processing or visuomotor speed were also included where they involved significant visual perceptual or visuomotor processes e.g. WAIS coding subtest.

2.2.2 Inclusion criteria for outcome measures

All types of functional outcome were included for review, with functional outcomes defined as any outcomes related to functioning in everyday life, including mobility, personal care, social skills, participation in the community and employment. Definitions of all functional outcomes included in this review are listed in the results section.

2.3 Data extraction

Data were extracted using a bespoke data extraction tool designed to best answer the research questions (See Appendix B).

2.3.1 Categorisation of cognitive assessments into domains

Assessments were initially categorised using the criteria listed above in section 2.2.2. Some test data, which were initially extracted for use in the domains of executive function or visuospatial perception/construction, were then extracted and analysed as a new domain labelled 'visuomotor speed' (VMS). This was done where visuomotor speed was a principal component of the task (e.g. WAIS coding subtest & TMT A) due to the significant involvement of visuomotor processes in many executive function and visuospatial

assessments, making this separation helpful in attempting to untangle the predictive value of these domains. 'Coding' subtests were then removed entirely from the domain of visuospatial perception/construction as research has shown that they principally measure processing speed and psychomotor processes (Kennedy et al., 2003). The TMT A was used in both EF and VMS domains as it involves both visuomotor and executive skills (Rossini & Karl, 1994).The TMT B was not used in the VMS domain as results are largely influenced by task-switching abilities (Arbuthnott & Frank, 2000), which are an executive skill.

2.3.2 Categorisation of outcome measures

Outcome measures were categorised based on the functional outcome that the scale was created and validated to measure, for example the Barthel Index (Mahoney & Barthel, 1965) was categorised as a measure of Activities of Daily Living (ADL) and therefore grouped with other ADL scales. This sometimes differed from the categorisation given by the experimental paper authors. Categories were further divided where papers split results for analysis (e.g. the Functional Independence Measure, a measure of ADL, is often split into cognitive and motor categories). Further break down of categories was not carried out due to few authors performing separate analyses for scale items.

2.4 Method of quality appraisal

Quality was assessed using a version of Downs and Black's (1998) quality index modified for use with non-intervention studies (Ferro & Speechley, 2009). The index was used to give each paper a numerical 'Quality Appraisal Rating', which was used to inform weighting and interpretation of data and any conclusions drawn. This quality index was chosen due to its use in the previous review (Mole & Demeyere, 2020) and other ABI research (Kinsella et al., 2015). Items are each scored dichotomously, as 1 (yes), or 0 (no/unable to determine), and a higher total score represents greater methodological quality.

The scale comprises four subscales: reporting (0–7), external validity (0–3), internal validity (0–4) and statistical power (0–1).

2.5 Data analysis

All data were first collated and summarised by domain in order to explore relationships between testing in specific cognitive domains and different functional outcomes. Study information and quality appraisal scores were used to inform conclusions. Data for specific cognitive assessments showing good evidence of predictive validity were then extracted and summarised. Good evidence of predictive validity was defined as being a significant independent predictor in 'well-controlled' multiple regression analyses or selection in best fit predictive models. Well-controlled regressions were defined as those controlling for demographic, medical and cognitive factors which had shown significant relationships in univariate analyses or were indicated as important in previous research. This decision was taken in order to increase the reliability of conclusions drawn from this section as their intended use is to inform clinical selection of cognitive assessment tools. Internal/external validity of studies was also assessed before this data was included due to the necessary reliance on statistical significance. Non-significant relationships between cognitive test scores and functional outcomes were also included, where these were provided, to reduce emphasis on null hypothesis significance testing (Cumming, 2014).

3 Results

The Electronic database search returned 10, 564 papers, with 17 of these meeting inclusion criteria for the current review. See Figure 1 for a summary of the study selection process. A summary of the design and methodology of all included studies is available in Appendix C.

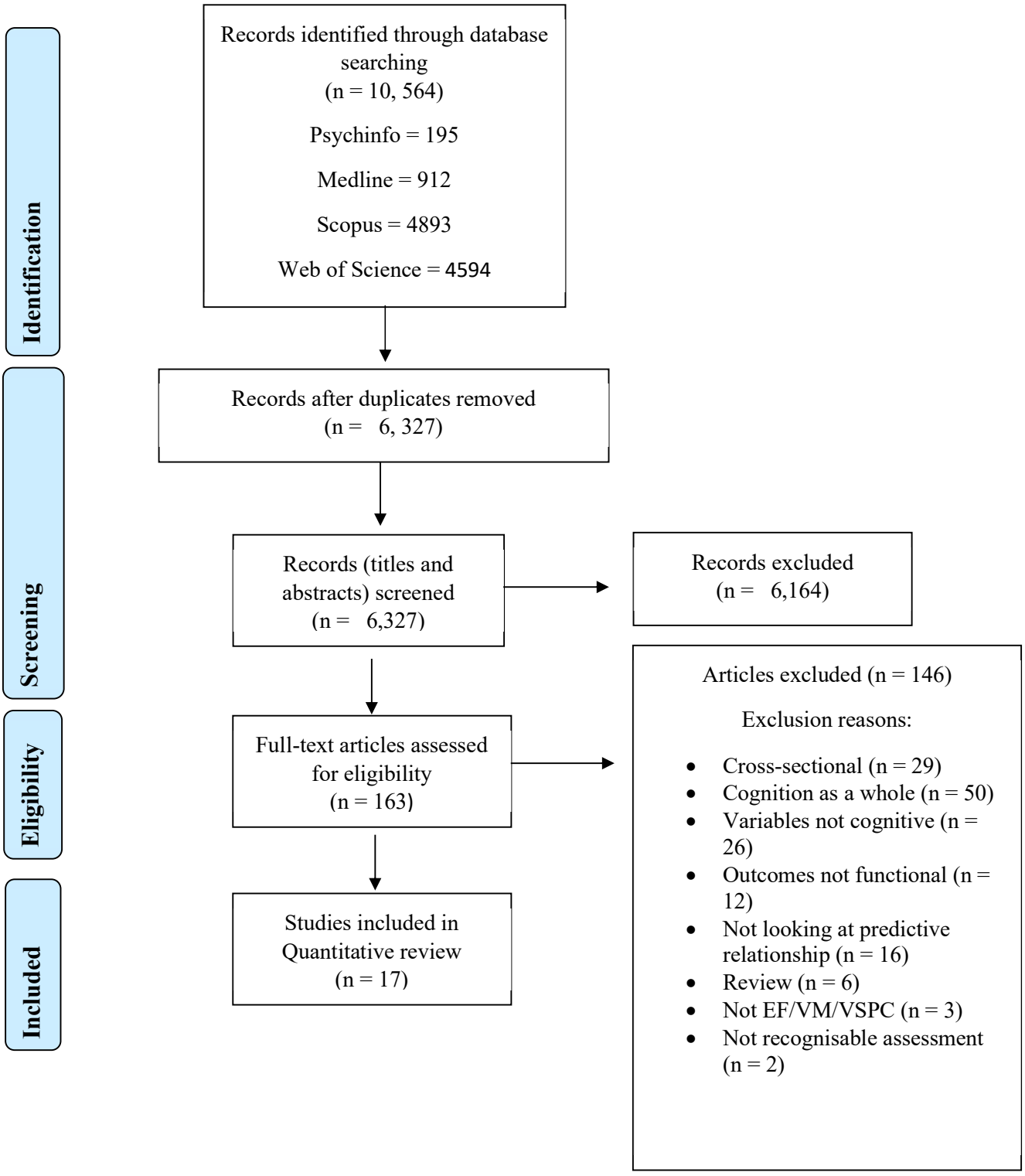

Figure 1: Study screening and selection process

3.1 Overview of studies

Studies were conducted in a range of countries including Israel, the United States, The United Kingdom, Australia, Canada, Finland, Singapore, The Netherlands, Korea, Norway and Denmark. The most common cognitive domain assessed was EF (13 studies; including one where data were extracted from the study author-categorised domain of attention) followed by VSPC (11 studies), VMS (six studies including two extracted from study author domains of attention and one from EF) and VM (four studies). Seven different categories of functional outcome were assessed, a summary of domains and outcome measures investigated is presented below in Table 1. Outcome categories were defined as:

- *Quality of Life* (QoL; seven studies); The quality of a person's health, comfort and happiness.
- *Activities of Daily Living* (ADL; eight studies): A person's ability to independently complete basic activities required for daily life, for example eating, walking and going to the bathroom. This category was further divided into two categories, 'motor' and 'cognitive' in one study.
- *Instrumental Activities of Daily Living* (IADL; three studies): A person's ability to complete activities allowing them to live independently in the community, though not required to support basic life, for example handling finances, using transport or doing household chores.
- *Participation* (three studies): This category was broken into three further categories; Participation in meaningful activities, participation in social relationships and participation in ADLs (defined as how often a person engages in ADLs rather than their ability to).
- *Reintegration* (two studies): Reintegration into normal recreation, movement in the community, and interaction in family or other relationships.

- *Handicap* (one study): Overall level of disability, including the following areas: Physical independence, cognitive independence, mobility, occupation, social integration, economic self-sufficiency.
- *Employment* (one study): Whether a person is employed in paid work.

Table 1: Summary of domains and outcomes investigated

Domain / Outcome measure	Executive function	Visuospatial perception/ construction	Visuomotor speed	Visual memory
QoL	4 studies (Adamit et al., 2015; Bertolin, etal., 2018; Cumming et al., 2014; Hochstenbach et al., 2001; Nys et al., 2006; Van Zandvoort et al., 2005)	5 studies (Cumming et al., 2014; Hochstenbach et al., 2001; Larson et al., 2003; Nys et al., 2006; Van Zandvoort et al., 2005)	2 studies (Hochstenbach et al., 2001; Van Zandvoort et al., 2005)	1 study (Nys et al., 2006)
ADL	5 studies (Laakso et al., 2019; Nys et al., 2005; Park et al., 2016; Park et al., 2017; Shea-Shumskye al., 2019)	6 studies (Larson et al., 2003; Narasimhalu et al., 2011; Nys et al., 2005; Park et al., 2016; Park et al., 2017; Wagle et al., 2011)	3 studies (Narasimhalu et al., 2011; Park et al., 2017; Wagle et al., 2011)	3 studies (Narasimhalu et al., 2011; Nys et al., 2005; Park et al., 2016; Park et al., 2017)
IADL	2 studies (Kapoor et al., 2019; Nys et al., 2005)	2 studies (Larson et al., 2003; Nys et al., 2005)	1 study (Kapoor et al., 2019)	1 study (Nys et al., 2005)
Handicap	No studies	1 study (Larson et al., 2003)	No studies	No studies

Reintegration	2 studies (Kapoor et al., 2019; Ownsworth & Shum, 2008)	No studies	1 study (Kapoor et al., 2019)	No studies
Employment	1 study (Ownsworth & Shum, 2008)	No studies	No studies	No studies
Participation in meaningful activities.	2 study (Adamit et al., 2015; Bertolin et al., 2018)	No studies	No studies	No studies
Participation (divided into social roles and ADLs)	1 study (Desrosiers et al., 2008)	1 study (Desrosiers et al., 2008)	No studies	No studies

3.2 Sample characteristics

All studies used samples originally recruited with consecutive opportunistic sampling methods, from hospital, rehabilitation or clinic settings, apart from (Park et al., 2017) who used a retrospective cohort sample from a stroke clinic. Overall, the samples were comprised of slightly more men than women, had a wide age-range of 19-98 and combined mean age of 65.15 (11.15) years. The most common type of stroke by far was ischemic, with approximately 80% of participants suffering an ischemic stroke, where stroke type was reported (13 out of 17 studies). Only two studies (Nys et al., 2005; Park et al., 2017) reported the prevalence of domain-specific impairments with a large variance in reported prevalence (EF: 29.7-60.5%, VM 16.2-59.6%, VSPC 30.6-62.5%). Participant characteristics for each study are displayed in Appendix D.

3.3 Quality appraisal of studies

A summary of individual scores on the modified Downs and Black (1998) quality index (Ferro & Speechley, 2009) is presented in appendix E, the maximum score on the index is 15.

3.3.1 Reporting

Reporting was generally rated highly with all studies clearly stating aims and outcome measures. However, six studies did not provide raw outcome data for the main findings, two did not provide information on the variability of their data and four did not report exact *p* values for findings. Six studies did not report the education level of their sample participants, despite this being a significant confound for performance on cognitive assessments. An additional problem with reporting was the general under-reporting of statistics for non-significant results and a lack of reporting of effect sizes for individual variables in regression modelling. These were not included as rating categories in the quality appraisal tool.

3.3.2 External Validity

All studies had inclusion/exclusion criteria which meant that samples were unlikely to be representative of their source populations. Nine studies excluded patients with previous psychiatric histories, comorbid medical conditions or neurological disorders and five excluded patients with neglect or aphasia. Several also imposed an upper age limit. Additionally, Narasimhalu et al., (2011) included only participants who has experienced a TIA or non-disabling ischemic stroke, meaning their sample was not generalisable. One study (Kapoor et al., 2019) sourced participants from a stroke prevention centre, meaning their samples may not have been ecologically valid. Four studies (Kapoor et al., 2019; Laakso et al., 2019; Narasimhalu et al., 2011; Shea-Shumsky et al., 2019) used samples and data from previous studies including a stroke prevention trial (Narasimhalu et al., 2011) and a rehabilitation treatment study (Shea-Shumsky et al., 2019). As these were treatment trials, they may have impacted on the relationships examined. Reporting of stroke characteristics showed that samples were mainly made up of participants with mild-moderate stroke meaning that findings will not be representative of patients with more severe impairments.

3.3.3 Internal validity

All studies used appropriate statistical analyses, though only nine controlled adequately for confounds and eight used stepwise or hierarchical regression analyses which can result in important variables being discarded depending on when they are entered into the model (Aaronson, 1989). All studies used validated outcome measures, although Van Zandvoort et al. (2005) observed ceiling effects in the BI and mRS. Many studies dichotomised data from continuous outcome scales without normed cut-off scores, which could compromise the validity of the categorised data. None of the seven studies using composites of author-chosen subtests reported internal validity statistics for these composites, making drawing conclusions from these results problematic. Without internal consistency statistics, we cannot be sure that the subtests were measuring the same domains or processes.

3.3.4 Statistical power

No study reported power analyses, and it is possible that studies were underpowered for multiple predictors and analyses. Where individual R squared statistics were reported it was possible to undertake post hoc power analyses, these are reported in the results section and used to aid in the interpretation of results.

3.4 Relationship between domain-specific impairment and later functional outcomes

3.4.1 Executive function (EF)

3.4.1.1 Activities of Daily Living (ADL)

Measures of EF were found to be predictive of ADL outcomes in multivariate regression analyses in three studies (Laakso et al., 2019; Park et al., 2017; Shea-Shumsky et al., 2019) and in a univariate regression analysis in one study (Nys et al., 2005). Effect sizes for the multiple regression analyses, variably reported as standardised beta, R squared and odds ratios, were fairly small, however this is generally expected in multiple regressions modelling

human outcomes. The interpretation and comparison of these effect sizes is statistically problematic, as discussed fully in the discussion section of this paper, therefore these statistics will not be overly relied upon for interpretation of results. Three other studies (Bertolin et al., 2018; Nys et al., 2005; Park et al., 2016) found relationships between measures of EF and ADL were not significant. However, Shea-Shumsky et al. (2019) had potential problems with the external validity of their sample and Bertolin et al. (2018) and Nys et al. (2005) both used stepwise or hierarchical regression analysis which has been found to have methodological issues.Park et al. (2016) used a verbal fluency subtest as their measure of EF, it may be that this aspect of executive function (known as initiation) is not related to later ADL outcomes as Park et al. (2017) also found that a measure of initiation was not a significant predictor of ADL outcomes. All results and available statistics are displayed in Table 2.

Table 2: Summary of findings for the predictive ability of executive function for activities of daily living outcomes

Study	Results	Quality appraisal rating
(Bertolin et al., 2018)	TMT A & B were not significant independent predictors of ADL (TMTA β = -.066; TMT B β= .025) in the hierarchical regression analysis. *p* values were not reported for non-significant results.	**10**
(Laakso et al., 2019)	Set shifting subdomain at baseline was significant predictor of mRS at 15 months regardless of stroke severity Model 1, *N* = 54, OR = 0.71, *p* = .018 Model 2, *N* = 54, OR = 0.75, *p* = .049 Other EF domains did not reach significance in model 1. Inhibition: *N* = 57, OR = 0.89, *p* = .133 Processing speed: *N* = 57, OR = 0.91, *p* = .434 Initiation: *N* = 57, OR = 0.75, *p* = .397 Strategy formation: *N* = 56, OR = 0.75, *p* = .466	**12**

(Park et al., 2016)	EF was not a significant predictor of either KBI (p = .18) or mRS (p = .353). Regression co-efficients were not reported for non-significant results.	**10**
(Shea-Shumsky et al., 2019)	BDS was a significant independent predictor of ADL scores at all four time points 3 months, N = 213, β = -0.245, p = .003 6 months, N =201, β = -0.257, p = .004 9 months, N = 188, β = -0.231, p = .016, 12 months, N = 188, β = -0.214, p = .006)	**8**
(Park et al., 2017)	t-test found participants categorized as having impaired EF scored significantly lower on KBI at discharge p < .05, no significant difference in mRS scores. Univariate linear regression showed TMT A & B, animal naming and digit-symbol coding EF scores at baseline were a significant predictor of KBI at discharge (effect sizes were not reported). Multivariate linear regression found that only TMT A was a significant independent predictor of KBI at outcome (N = 104, β= 0.288, p = .008. Other subtests were not significant independent predictors: TMT B: N = 104, β = -0.186, p = .071 Digit-symbol coding: N = 104, β = -0.055, p = .071 Animal-naming: N = 104, β = -0.064, p = .482	**11**
(Nys et al., 2005)	EF at baseline was significantly associated with BI at follow-up in the univariate analysis (p < .01). EF was not a significant independent predictor in either of the regression models. Non-significant statistics were not reported.	**12**

3.4.1.2 Quality of Life (QoL)

Measures of EF were found to be significant predictors of QoL outcomes in one well-controlled multivariate and one univariate regression (Hochstenbach et al., 2001; Nys et al., 2006). However, Cumming et al. (2014) found that the relationship did not reach significance in a univariate regression and Nys et al. (2006) found that EF did not reach significance in a

multivariate regression analysis. As above, differences in type of cognitive assessment used could account for this. Cumming et al.'s (2014) composite contained two language initiation subtests which were found to have no significant relationship with other functional outcomes (e.g. Laakso et al., 2019). Issues surrounding the use of stepwise regression and potential lack of power may account for Nys et al.'s (2006) non-significant result. As statistics were not reported for these results, we cannot assess the size/nature of any possible effect, beyond its non-significance at $p < .05$. Results are summarised in Table 3.

Table 3: Summary of findings for the predictive ability of executive function assessments for quality of life outcomes

Study	Results	Quality appraisal rating
(Cumming et al., 2014)	EF as a domain did not reach Bonferroni adjusted significance in univariate linear regression (β = 0.04, 95% CI −0.02–0.10, $p = .192$).	**8**
(Nys et al., 2006)	EF at baseline was a significant predictor of QoL at 6 months in univariate regression (β =0.21; $p < .05$). EF was not a significant predictor in multiple stepwise regression analysis, statistics were not reported.	**12**
(Hochstenbach et al., 2001)	TMT A & B were significantly related in the Univariate ANOVA and chi-squared test to QoL scores at follow-up. TMT B was the only cognitive variable and one of only two variables (gender being the other) that was a significant predictor in regression analyses (LRS = 11; df = 3; $p = .012$. TMT A was not a significant predictor (statistics for non-significant results were not reported).	**10**

3.4.1.3 Participation

Adamit et al. (2015) and Desrosiers et al. (2008) found that measures of EF were independent predictors of participation in meaningful activities and social relationships in well-controlled multiple regression analyses. Desrosiers et al. (2008) also found that EF

measures were selected for a best fit model of participation in ADLs but this relationship did not reach significance. Post-hoc power analyses indicated that this analysis did not have sufficient power to detect a significant relationship for this effect size, if one exists. This could explain the non-significant result. However, the effect size is extremely small indicating that this relationship, if present, is not strong. Bertolin et al. (2018) however found no significant relationship between their measure of EF and participation and Adamit et al. (2015) found significant relationship for only one of their EF assessment tools. As above, this may be due to different assessment tools as significant relationships were found for measures of inhibition, whereas Bertolin et al. (2018) used measures of planning and set-shifting. Results are summarised in Table 4.

Table 4: Summary of findings for the predictive ability of executive function assessment for participation outcomes

Study	Results	Quality appraisal rating
(Adamit et al., 2015)	DEX was an individually significant predictor of Participation at 6 months. EFPT was not a significant predictor. DEX β = 0.265, *p* = .014; EFPT β = -0.087, *p* = .183.	**11**
(Bertolin et al., 2018)	TMT A & B were not significant independent predictors of participation (TMT A β = -.018; TMT B β = .005*), p* values were not reported.	**10**
(Desrosiers et al., 2008)	EF at baseline was not part of the best fit models for predicting participation outcomes for the acute care group. EF was in best model for predicting participation in both social roles and ADLs for the rehabilitation sample. EF was a significant independent predictor of participation in social roles (*N* = 95, β = -0.114, *p* = .002, change in R^2= 0.14). Post-hoc power calculations completed on G*power, with error set at .05, demonstrated a very good power of 0.97 indicating that the model had sufficient power to detect this relationship. EF was not a significant independent predictor of participation in ADLs (*N* = 91, β = -0.052, *p* = .48, change in R^2 = .02)	**11**

	Post-hoc power calculations on G*power with the alpha set at .05 showed that the study had a power of 0.27 indicating it was underpowered to detect a significant relationship with this effect size.	

3.4.1.4 Instrumental Activities of Daily Living (IADL)

Two studies (Kapoor et al., 2019; Nys et al., 2005) found univariate relationships between EF and IADL with small to medium effect sizes. However, potential confounds were not controlled for and Nys et al. (2005) found this relationship did not remain significant in multivariate regression analysis. Results are summarised in Table 5.

Table 5: Summary of findings for the predictive ability of executive function assessment for instrumental activities of daily living outcomes

Study	Results	Quality appraisal rating
(Kapoor et al., 2019)	TMT A and B both significantly correlated with FAI scores (r = 0.37, p = .016; r = 0.54, p < .001).	**10**
(Nys et al., 2005)	EF at baseline was significantly associated with FAI scores at follow-up in the Univariate analysis (p < .01). EF was not a significant independent predictor in either of regression model. Statistics were not reported for non-significant results.	**12**

3.4.1.5 Reintegration

Ownsworth and Shum (2008) found that some measures of EF correlated with later reintegration scores with small to medium effect sizes, whereas Kapoor et al. (2019) found there were no significant correlations between their measures of EF and reintegration. Results are summarised in Table 6.

Table 6: Summary of findings for the predictive ability of executive function assessment for reintegration outcomes

Study	Results	Quality appraisal rating
(Kapoor et al., 2019)	TMT A (r = -.153, p = .330, n = 44) and B (r = -.058, p = .713, n = 44) did not significantly correlate with RNLI scores	**10**
(Ownsworth & Shum, 2008)	Five point test-designs (r = 0.48, p < 0.05), FAS-errors (r = -0.61, p <0.01) and Tinker Toy test (r = 0.38, p <0.05) scores at baseline were significantly correlated with reintegration scores at outcomes. Other EF tests were not significantly correlated. Health and safety subtest (r = 0.24) Key search (r = 0.30)	**10**

3.4.1.6 Employment

Measures of EF at two years post-stroke were found to be able to discriminate between employment outcomes at over three years post-stroke in univariate analysis (Ownsworth & Shum, 2008). However, no other confounds than time were controlled for and sample sizes were small (only 10 unemployed participants). Results are summarised in Table 7.

Table 7: Summary of findings for the predictive ability of executive function assessment for employment outcomes

Study	Results	Quality appraisal rating
(Ownsworth & Shum, 2008)	Five-Point Test total designs (t = -2.29, p = .037) and the Tinker Toy test (t = -3.66, p < .001) scores at baseline significantly discriminated between employed and unemployed groups at follow-up. Other measures of EF were not significant. Health and safety subtest: t = -0.46, p = .65 Key search test: t = -1.81, p = 0.082 FAS test: t = -0.22, p = 0.826	**10**

3.4.2 Visuospatial perception/construction (VSPC)

3.4.2.1 Activities of Daily Living (ADL)

Three studies found measures of VSPC were significant independent predictors of ADL outcomes in well-controlled multivariate regression analyses (Larson et al., 2003; Park

et al., 2016; Wagle et al., 2011) all with substantial effect sizes indicating strong relationships. Three others found significant relationships in univariate analyses but these did not reach significance in multivariate regression (Narasimhalu et al., 2011; Nys et al., 2005; Park et al., 2017). Factors that could account for these non-significant results include the use of different assessment tools for this cognitive domain. Narasimhalu et al.'s (2011) sample may also have lacked external validity as it only contained people who had suffered a TIA or non-disabling ischemic stroke. Nys et al. (2005) used a stepwise regression, issues with this are discussed above. Results are summarised in

Table **8**.

Table 8: Summary of findings for the predictive ability of visuospatial perception/construction for activities of daily living outcomes

Study	Results	Quality appraisal rating
(Narasimhalu et al., 2011)	Visuoconstruction (HR = 3.52, $p < .001$) was a significant predictor of mRS scores in the univariate analysis. It was not a significant independent predictor in the Multivariate regression. (HR = 1.18, $p < .215$).	**9**
(Park et al., 2016)	Linear regression showed CPT ($R^2 = 0.496$, $p = .044$) was an independent predictor for functional improvement measured by the change in K-MBI during the first 3 months after stroke. A post-hoc power calculation could not be undertaken due to no *N* being reported for the analysis. Stepwise logistic regression showed that the z-score of CPT (OR= 0.282, $p = .049$) was also independent predictor of mRS scores at 6 months.	**10**
(Wagle et al., 2011)	VCI ($\beta = 0.459$, $p < .001$) and all individual VCI subtests were significant in the unadjusted regression. In the adjusted regression VCI was the only RBANS index which remained significant ($\beta = –0.309$, $p < .001$). Figure copy (VCI index) was one of only two subtests which remained significant independent predictors in the adjusted multiple regression ($\beta = 0.233$, $p = .002$).	**12**
(Park et al., 2017)	*t*-test showed people with VS impairment scored significantly lower on KBI ($p < .05$) at discharge (no significant difference in mRS scores).	**11**

	ROCF was a significant predictor of mRS in univariate regression. (statistics not reported). ROCF was not a significant independent predictor in the multiple regression (β = 0.115, *p* = 0.271).	
(Nys et al., 2005)	Univariate analysis found that impairment in perception/construction significantly associated with poorer BI scores at follow-up. *p* < .01. VPC was not a significant predictor of BI in either regression model. Statistics for non-significant results were not reported.	**12**
(Van Zandvoort et al., 2005)	Ceiling effects found for BI and mRS	**4**
(Larson et al., 2003)	Cognitive Functioning: VCI significantly correlated but did not reach Bonferroni significance. VCI was a significant independent predictor in multiple regression (β = .31). Motor Functioning: VCI significantly correlated but did not reach Bonferroni significance. VCI was the only significant independent predictor in multiple linear regression (β = .51).	**7**

3.4.2.2 Quality of Life (QoL)

Two studies (Cumming et al., 2014; Nys et al., 2006) found measures of VSPC were significant independent predictors of QoL in well-controlled multivariate regression analyses, with Nys et al. (2006) finding it was the single strongest predictor. Reported effect sizes for Cumming et al. (2014) however were very small indicating the amount of variance predicted may have been negligible. Difficulties with interpreting this are discussed in later sections. One study (Hochstenbach et al., 2001) found significant univariate relationships but no significant relationship in a stepwise regression analysis. As above, this could be due to either differences in tests assessed, or the use of stepwise regression analyses with multiple

variables and a fairly small sample size of 106. Van Zaandvort et al. (2005) found no significant correlations between these two variables, however this study was compromised by methodological issues. A summary of results is displayed in Table 9.

Table 9: Summary of findings for the predictive ability of visuospatial perception/construction assessment for quality of life outcomes

Study	Results	Quality appraisal rating
(Cumming et al., 2014)	Univariate linear regressions showed 3- month visuospatial ability (β = 0.10, 95% CI 0.05–0.15, $p < .001$) significantly associated with 12-month AQoL. Visuospatial ability significant independent predictor in multivariate regression (β = 0.08, 95% CI 0.03–0.12, $p = .002$; $R^2 = 0.37$).	**8**
(Nys et al., 2006)	VPC was significant predictor of QoL in Univariate regression (β = − 0.47; $p < .001$) In multiple stepwise analysis, an impairment in visual perception/construction was the strongest risk factor of a reduced QOL (β = − 0.44; $p < .001$). Post hoc analyses revealed that patients with a deficit in visual perception/construction reported the most complaints in terms of the number of affected SS-QOL domains.	**12**
(Van Zandvoort et al., 2005)	No significant correlation between VCI and QoL. Statistics not reported.	**4**
(Hochstenbach et al., 2001)	Univariate ANOVA found that WAIS block design at baseline were significantly related to QoL at follow-up (F = 5.70, $p = .08$) WAIS block design (with time limit) and Money's Road map were near significance on Chi-Squared test ($p = .051$; $p = .082$) and were included in the multiple regression. No tests from VCP domain were significant predictors in stepwise multiple regression (non-significant results were not reported).	**10**

3.4.2.3 Instrumental Activities of Daily Living (IADL)

Strong evidence for the predictive relationship between VSPC and IADL was found from two multivariate regression analyses (Nys et al., 2005), one of which was classified as well-controlled. Results are summarised in Table 10.

Table 10: Summary of findings for the predictive ability of visuospatial perception/construction assessment for instrumental activities of daily living outcomes

Study	Results	Quality appraisal rating
(Nys et al., 2005)	Univariate analysis found that impairment in perception/construction significantly associated with poorer IADL scores at follow-up ($p < .01$). Stepwise multiple regression found that only visual perception and construction remained a significant predictor for IADL. In stepwise multiple regression with other cognitive and medical variables Visual perception/construction was still a significant predictor with impairment in this domain at baseline associated with 27 times greater odds of impairment in IADL at follow-up (other statistics were not reported).	**12**
(Larson et al., 2003)	VCI positively correlated but did not reach Bonferroni significance. Multiple regression found that VCI was the only significant independent predictor ($\beta = .47$).	**7**

3.4.2.4 Participation

Evidence of a link between early VSPC and later participation in social roles was found, with VSPC test scores selected for the best fit model but not reaching significance as an independent predictor (Desrosiers et al., 2008). The lack of significance as an independent predictor may be due to insufficient power, as post-hoc power calculations indicated the study was underpowered to detect a significant relationship. It could also have been impacted by the order of variables selected in stepwise analysis. Results are summarised in Table 11.

Table 11: Summary of findings for the predictive ability of visuospatial perception/construction assessment and later participation outcomes

Study	Results	Quality appraisal rating
(Desrosiers et al., 2008)	VPC was in the best fit model for predicting social participation in the Rehabilitation sample. VPC was not a significant independent predictor. $\beta = 0.088$ R^2 change = 0.04 $p = .004$ A post-hoc power calculation completed using G*power, with alpha set at 0.05, calculated power to be 0.5 for this analysis. This indicated that the analysis did not have sufficient power to detect a significant relationship for this effect size.	**11**

3.4.2.5 Handicap

Evidence from a multiple regression analysis indicated that VSPC is a significant independent predictor of later handicap (Larson et al., 2003). While Larson et al.'s (2003) quality appraisal score was low, this was mainly due to a lack of reporting rather than issues with internal/external validity. However, potential confounds (other than additional cognitive domains) were not controlled for. Results are summarised in Table 12.

Table 12: Summary of results for the predictive ability of visuospatial perception/construction assessment for handicap outcomes

Study	Results	Quality appraisal rating
(Larson et al., 2003)	Handicap: VCI significantly correlated but did not reach Bonferroni significance. VCI only significant independent predictor (β = .38) in multiple regression.	**7**

3.4.3 Visual Memory

3.4.3.1 Activities of Daily Living (ADL)

Measures of VM were found to be significant predictors of later ADL in two well-controlled multivariate regression analyses (Nys et al., 2005; Park et al., 2016). Narasimhalu et al., (2011) found a significant relationship in univariate analyses, but this was not significant in a multivariate regression. As stated above this could be due to problems with the external validity of the sample used, which may have prevented this relationship from reaching significance in multivariate analyses. Results are summarised in Table 13.

Table 13: Summary of findings for the predictive ability of visual memory assessment for activities of daily living outcomes

Study	Results	Quality appraisal rating
(Narasimhalu et al., 2011)	VM was significant predictor in Univariate analysis (HR = 2.4 p = .014) but not in the multivariate (HR = 1.05, p = .870).	**9**
(Park et al., 2016)	VM was an independent predictor of change in KBI in the multiple linear regression (R^2 =0.614, p < .001). A power calculation could not be completed due to N not being reported for individual analyses. It was also a significant predictor of mRS in the stepwise logistic regression (OR= 0.529, p = .048).	**10**
(Nys et al., 2005)	VM was significantly associated with scores on BI in Univariate analysis (p < .01). VM was the only cognitive domain that was a significant predictor of BI score in the stepwise logistic regression and remained significant when medical predictors were added. Patients with baseline deficits in visual memory had 22 times greater odds of having impairments in ADL at follow-up (other statistics were not reported).	**12**

3.4.3.2 Quality of Life (QoL)

VM was found to be a significant predictor of later QoL in univariate regression analysis but this relationship did not reach significance as in independent predictor in multiple stepwise analysis (Nys et al., 2006). The fact that this relationship did not reach significance in multivariate regression could be due to a lack of power or the order variables entered the stepwise regression. Results are summarised in Table 14.

Table 14: Summary of findings for the predictive ability of visual memory assessment for quality of life outcomes

Study	Results	Quality appraisal rating
(Nys et al., 2006)	VM was significant predictor in the Univariate regression (β = − 0.37; p < .001) but not in the multiple regression (non-significant statistics not reported).	**12**

3.4.3.3 Instrumental Activities of Daily Living (IADL)

VM was found to be significantly related to later IADL outcomes in univariate analysis but was not a significant independent predictor in multivariate regression (Nys et al., 2005). Again, this could be due to issues with power and the use of stepwise regression analysis. Results are summarised in Table 15.

Table 15: Summary of findings for the predictive ability of visual memory assessment for instrumental activities of daily living outcomes

Study	Results	Quality appraisal rating
(Nys et al., 2005)	VM was significantly associated with scores on FAI in Univariate analysis ($p < .01$). VM was not a significant independent predictor of IADL in the logistic regressions (non-significant statistics were not reported).	**12**

3.4.4 Visuomotor speed (VMS)

3.4.4.1 Activities of Daily Living (ADL)

Strong evidence for the predictive relationship between VMS impairments post-stroke and later ADL outcomes, up to 5 years later, with measures of VMS found to be significant independent predictors in three separate multivariate regression analyses (Narasimhalu et al., 2011; Park et al., 2017; Wagle et al., 2011), two of which were classed as well-controlled. Results are summarised in Table 16.

Table 16: Summary of findings for the predictive ability of visuomotor speed and activities of daily living outcomes

Study	Results	Quality appraisal rating
(Narasimhalu et al., 2011)	VMS (HR = 6.78, $p < .001$) was a significant predictor of mRS scores in the Univariate analysis. VMS was the only significant independent predictor in the multivariate analysis (HR=3.49, $p = .002$).	**9**

(Wagle et al., 2011)	Coding was one of only two subtests that remained significant in the adjusted multiple regression (β = –0.484, $p < .001$)	**12**
(Park et al., 2017)	Digit-symbol coding and TMT A were significant predictors of mRS in Univariate regression. Coding was not a significant predictor in the multiple regression. TMT A was a significant independent predictor of KBI at outcome (N = 104, β = 0.288, p = .008)	**11**
(Van Zandvoort et al., 2005)	Ceiling effects found for BI and mRS	**4**

3.4.4.2 Quality of Life (QoL)

A significant relationship was found between VM and QoL in one univariate analysis, however this relationship did not reach significance in a multivariate stepwise regression (Hochstenbach et al., 2001). As above, lack of significance could be due to the use of stepwise regression analyses which can influence significance of variables. Van Zandvoort et al. (2005) found no significant correlation between the two variables, however methodological issues may have impacted these findings (see quality appraisal). Results are summarised in Table 17.

Table 17: Summary of findings for the predictive ability of visual memory assessment and quality of life outcomes

Study	**Results**	**Quality appraisal rating**
(Van Zandvoort et al., 2005)	No significant correlation between VM and QoL (statistics were not reported).	**4**
(Hochstenbach et al., 2001)	Univariate ANOVA showed that TMT-A time (F = 5.02, N = 75) and coding (F = 9.25, N = 72) were significantly related to QoL at follow-up. Neither were found to be significant in the stepwise regression analysis (non-significant results not reported).	**10**

3.4.4.3 Instrumental Activities of Daily Living (IADL) and Reintegration

Some measures of VMS were significantly correlated with IADL outcomes, though effect sizes were small. but not reintegration scores in one study (Kapoor et al., 2019), however no confounds were controlled for. Evidence in summarised in Table 18.

Table 18: Summary of results for the predictive ability of visual memory assessment for instrumental activities of daily living and reintegration outcomes

Study	Results	Quality appraisal rating
(Kapoor et al., 2019)	Digit symbol coding was not significantly correlated with RNLI (R = -.189, *p* = .238) or IADL (R = .316, *p* = .050) scores. TMT A was significantly correlated with FAI scores (R = 0.37, *p* = .016) TMT A did not significantly correlate with RNLI scores (r = -.153, *p* = .320).	**10**

3.5 Predictive validity of neuropsychological subtests and composites

Eight subtests and four composites demonstrated evidence of predictive validity in well-controlled[2] multiple regression analyses. Relationships between cognitive tests and ADLs were by far the most studied and therefore resulted in the most significant relationships. The TMT A was found to be a significant independent predictor of ADL but not QoL, while the TMT B had the opposite relationship with these variables (Hochstenbach et al., 2001; Park et al., 2017). The difference between these two subtests is the set-shifting/task-switching element, indicating that impairments in this process may be related to a QoL outcomes, while visual attention, psychomotor and planning processes may be more important for predicting functioning in ADLs. However, the TMT A was found not to be predictive of ADLs in a separate study (Bertolin et al., 2018), potential reasons for this are discussed in the domain results section. Tests of inhibition (Stroop and DEX) showed evidence of predictive validity

[2] Well -controlled regression analyses were defined as those containing multiple demographic, medical and cognitive variables in order to control for possible confounds

for participation outcomes, particularly participation in social roles. Other tests showing predictive validity for ADLs involved visual attention, visuospatial construction and visuomotor speed. IADL and QoL were also predicted by visuospatial composites (Cumming et al., 2014; Nys et al., 2005) indicating that visual skills may also be important for these outcomes. Summary of all results for cognitive assessments showing predictive validity in multiple regression analyses are displayed in Table 19.

Table 19: Domain-specific cognitive assessment tools showing predictive validity for functional outcomes

Test	Skills/subdomains involved	Relationships assessed	Summary of evidence
DEX	Behavioural difficulties such as impulsivity, inhibition control, monitoring and planning.	Participation* – One study	Significant independent predictor of participation (β = 0.265) in well-controlled multiple regression analysis (Adamit et al., 2015).
TMT A	Visual attention, visual search (planning and speed), visuomotor speed.	ADL* – Two studies Participation – One study QoL – One study	Not a significant independent predictor of ADL (β = -.066) or Participation (β = -.018) in hierarchical regression analysis (Bertolin et al., 2018). Significant independent predictor of ADL (β = 0.288) in multiple linear regression (Park et al., 2017). Not a significant independent predictor of QoL in forwards-backwards stepwise regression (Hochstenbach et al., 2001), statistics not reported.
TMT B	Task-switching, visual search (planning and speed), visuomotor speed, visual attention.	ADL – Two studies Participation – One study QoL* – One study	Not a significant independent predictor of ADL (β = .025) or Participation (β = .005) in hierarchical regression analysis (Bertolin et al., 2018). Not a significant independent predictor of ADL (β = -0.186, *p* =

			0.071) in multiple linear regression (Park et al., 2017). Significant independent predictor of QoL (LRS = 11; df = 3; *p* = .012) in forwards-backwards stepwise regression (Hochstenbach et al., 2001)
TMT B and WCST	Task-switching, visual search (planning and speed), motor speed, visual attention.	ADL* – One study	Significant independent predictor of ADL (OR = 0.75) in multiple logistic regression (Laakso et al., 2019).
CRT	Visual attention, visuospatial perception and construction, visual memory, visuomotor skills	ADL* - One study	Significant independent predictor of ADL in multiple linear regression (R^2 = 0.614) and stepwise logistic regression (OR = 0.529). (Park et al., 2016).
CPT	Visual attention, visuospatial perception and organization, visuomotor skills.	ADL* – One study	Significant independent predictor of ADL in multiple linear regression (R^2 = 0.496) and stepwise logistic regression (OR = 0.282). (Park et al., 2016).
BDS	Behavioural regulation, inhibition.	ADL* – One study	Significant independent predictor of ADL at four timepoints (R^2 = 0.33, 0.276, 0.225, 0.348) in stepwise regression analysis (Shea-Shumsky et al., 2019).
Visual Memory composite: Corsi Block span, the ROCF- delay, the WMS–Visual Reproduction, and the Location Learning Task.	Visuospatial working memory, visual attention, visuospatial perception and construction, visuomotor skills.	ADL* – One study IADL – One study	Significant independent predictor of ADL (22 times greater odds of impairment) but not IADL in multiple logistic regression (Nys et al., 2005).
Visuomotor composite: Digit symbol coding, digit modality test, digit cancellation, and maze task.	Visual attention, processing speed, visual working memory, visuomotor speed, visuospatial	ADL* – One study	Significant independent predictor of ADL (HR = 3.49) in Cox Proportional Hazards regression (Narasimhalu et al., 2011).

	perception, visual planning.		
VPT vertical	Visual perception and visual attention.	Participation in social roles and ADLs – One study	Not a significant independent predictor for participation in stepwise multiple regression analysis (β = 0.088). However, was selected for inclusion in the best fit model (R^2 change = 0.04) for predicting participation in social activities (Desrosiers et al., 2008).
Stroop	Inhibition, processing speed	Participation in social roles* and ADLs – One study ADL – One study	Significant independent predictor of participation in social roles (β = - 0.114) in stepwise multiple regression. Selected for inclusion in the best fit model for participation in ADLs (β = -0.052) but not a significant independent predictor (Desrosiers et al., 2008). Not a significant predictor of ADLs in logistic regression (Laakso et al., 2019)
Visuospatial composite: Judgment of Line Orientation (short form), Test of Facial Recognition (short form), the Rey-Osterrieth Complex Figure–copy, and the WAIS-III block design.	Visuospatial perception, visuospatial construction, visual attention, visuomotor skills.	IADL* – One study ADL – One study QoL* – One study	Significant independent predictor of IADL but not ADL in stepwise logistic regression (Nys et al., 2005). Significant independent predictor of QoL (β = -0.44) in stepwise logistic regression (Nys et al., 2006).
ROCF-Copy and WAIS-R Block Design	Visual perception, visual attention, spatial processing, visuomotor skills	*QoL – One study	Significant independent predictor of QoL in multivariate regression (B = 0.08, 95% CI 0.03–0.12, $p = 0.002$; $R^2 = 0.37$) (Cumming et al., 2014).

4 Discussion

Testing in the cognitive domains of executive function (EF), visual memory (VM) and visuospatial perception/construction (VSPC), within 6-weeks post-stroke, has been found to be predictive of later functional 'activities' between 6- months -1 year later (Mole & Demeyere, 2020). 'Activities' were defined as including "learning and applying knowledge," "general tasks and demands," "communication," "mobility," and "self-care", and amalgamated results from scales measuring a variety of functional constructs, including Activities of Daily Living (ADL) and Quality of Life (QoL). The current review aimed to expand on these results in three main ways 1) to increase the breadth and generalisability of results by expanding inclusion criteria to include multiple assessment time-points relevant to different clinical services; 2) To separate and explore relationships between domain-specific tests and individual functional constructs (e.g. QoL, ADL) and to collate evidence for functional outcomes not yet reviewed, such as employment, Instrumental Activities of Daily Living, and Reintegration into the community; 3) To identify individual, domain-specific, clinical assessment tools that have predictive validity for later functional outcomes post-stroke. In order to further examine relationships between these domains and functional outcomes, a third cognitive domain defined as 'visuomotor speed' was extracted from EF and VSPC domains and examined separately. This was due to visuomotor processes often being a main component of tests of EF or VSPC and therefore potentially responsible for the predictive validity of assessment in these domains.

A larger quantity of studies was reviewed for each domain, compared to the previous review (Mole & Demeyere, 2020), with 13 studies of EF, 11 of VSPC, 6 VMS and 4 of VM. In general, quality of studies was high with most scoring ≥10 on the quality appraisal tool. There was, however, an underreporting of statistics for non-significant results, which was not captured by the quality appraisal tool, and a lack of population-based samples. There was also

a general over-reliance on statistical significance, especially considering that most studies did not report power calculations and were likely underpowered.

4.1 Relationships between domain-specific assessment and later functional outcomes

4.1.1 Generalisability to different assessment timepoints

Collated evidence suggests that the predictive relationship between domain-specific testing post-stroke and later functional outcomes is generalisable to a broad range of assessment time-points. Predictive relationships were found in well-controlled multivariate regression analyses for functional outcomes assessed at both at 3 months (Shea-Shumsky et al., 2019) and over 5 years post-stroke (Narasimhalu et al., 2011), suggesting that domain-specific impairments predict long-term functional recovery. Additionally, Shea-Shumsky et al. (2019) tested outcomes at four separate time-points (3,6,9 and 12 months) and found that this did not impact on effect size. Significant relationships were also found when cognitive assessments took place both early after stroke (e.g. Nys et al., 2006; Less than 3 weeks) and at later time-points (e.g. Ownsworth & Shum, 2008; Mean 2.1 years post-stroke) indicating that domain-specific cognitive assessments have clinical validity at later time-points, for example, in community neuropsychology services. The lack of any obvious differences in results for studies using different assessment timepoints could indicate that the timing of cognitive assessments doesn't impact their predictive ability, however, without more research this cannot be known as no study carried out the same cognitive testing at multiple time-points post-stroke.

4.1.2 Relationships between individual domains and functional outcomes.

Collated evidence indicates that the previous relationships found between testing in the cognitive domains of EF, VSPC,VM and later functioning in 'activities' (Mole & Demeyere, 2020), are also in evidence when the outcome variables are divided into individual functional constructs such as ADL and quality of life. The current review also demonstrates that this may be generalisable to other functional outcomes, such as participation in meaningful activities. The separated domain of VMS was also found to be predictive of functional outcomes. There was a large variance in the number of studies exploring each relationship, with some domains and outcomes investigated in multiple studies and some not at all.. Additionally, effect sizes from multiple regression analyses could not be reliably compared or interpreted as discussed more fully below. This made it difficult to compare relationships between different outcomes and cognitive variables, although there was some limited evidence that relationships differed in extent. For example, Nys et al. (2005) found that their composite of tests measuring VM was predictive of later ADL but not IADL outcomes, however as effect sizes were not reported it cannot be known if these relationships truly differed, except for in significance. Most of the relationships found to be significant in one or more studies did not reach significance in other studies and problems with the reliance on statistical significance were compounded by lack of power analyses. Evidence was therefore collated and evaluated with conclusions informed by quality appraisal of evidence.

At least some of the differences in the significance of findings may be attributable to the tool used to assess the cognitive domain, although again, this cannot be known as studies using different tests also used different participants, testing time-points and outcome measures, any/all of which may have influenced results. There was some evidence supporting the hypothesis that the individual assessment tool used impacted on the predictive

relationship. Laakso et al. (2019) found that only one of their five measures of EF was a significant predictor of ADL outcomes despite them all being used with the same population, testing timepoint and outcome measure. This provides support for the utility of providing evidence of the predictive value of individual assessment tools (discussed below).

Overall, functioning in the domain of EF was found to be predictive of the following functional outcomes: ADL, QoL and participation in meaningful activities. The domain of VSPC was found to be predictive of ADL, QoL, IADL and participation in social relationships. The domains of VM and VMS were both found to be predictive of later functioning in ADLs. The fewer relationships found for VM and VMS is likely due to a lack of investigation, as other relationships were indicated for these domains, but not enough evidence was available to form conclusions. It is important to keep in mind that conclusions could be dependent on sample, time-point or cognitive assessment tool used, and so only infer predictive ability of these domains in certain circumstances.

4.2 Predictive ability of individual assessment tools

Overall, eight subtests and two composites showed good evidence of predictive validity for later functional outcomes post-stroke. These are displayed in Table 19 along with corresponding evidence. There was some evidence that the different subdomains/cognitive processes involved in tasks accounted for differences in predictive validity. Assessment tools involving language initiation in the domain of EF were not found to be predictive in any study whereas tests involving set-shifting and inhibition were. There was also evidence that the subdomains, or combination of domains, assessed by subtests/composites were differentially predictive of outcomes. For example, assessments primarily involving the executive function subdomain of inhibition were found to be linked to participation in social relationships whereas tests involving general visual processes (including visuospatial,

visuomotor and attention) were linked to later ADL outcomes. These results indicate that the predictive validity of individual assessment tools is likely to be more clinically useful information than the predictive ability of domain-specific testing, which could be influenced by the combination of cognitive domains and processes involved in tests.

4.3 Strengths and limitations of review

A strength of this review is the broad inclusion criteria, resulting in a relatively large number of papers being identified in comparison to previous reviews. At least four papers were reviewed for each cognitive domain, increasing validity and generalisability of conclusions. Studies came from a wide range of countries and included participants with a large age range and multiple types of stroke. However, generalisability may have been compromised by the fact that most research papers had strict inclusion/exclusion criteria for their participants with most excluding patients with more severe impairments or co-morbid medical or psychological conditions.

Other limitations were related to the variance in number of studies investigating each cognitive domain and outcome, and the difficulties with comparing effect sizes. Many significant relationships were claimed where effect sizes appeared extremely small (e.g. Park et al., 2017), however effect sizes from multiple regression models are highly context dependent and small effect sizes do not necessarily invalidate a significant relationship, particularly in human behaviour/cognition studies where a high variance in results is expected. Research generally shows that effect sizes such as beta values, R squared and odds/hazard ratios cannot be reliably interpreted or compared across studies due to their relationship to sample size, measures/variables studied and variance in results (e.g. Uanhoro et al., 2019). This makes it necessary to instead rely on the significance of results, which

presents its own difficulties as discussed below, meaning that conclusions around differential relationships between variables could not be formed.

4.4 Clinical implications and recommendations for future research

This review adds further evidence to the predictive ability of assessment in the domains of EF, VSPC, VM and VMS, post-stroke for later functional outcomes. These relationships exist at later assessment time-points and are predictive of long-term future outcomes in some circumstances. The extent of the predictive ability of cognitive assessments in these domains for functional outcomes were not determined due to difficulties with the interpretation of effect size: It could be that the relatively small effect sizes found in many instances indicate that cognitive assessment post-injury has limited utility for predictive purposes. It is suggested that future research calculates and reports statistics that mitigate some of the problems presented, such as risk ratios. There was evidence that the predictive nature of domains may be a consequence of specific subdomains or combination of domains assessed by individual cognitive tools. It is therefore recommended that evidence regarding the predictive validity of specific assessment tools, rather than of general testing in specific cognitive domains, is utilised for clinical purposes. More research untangling cognitive assessment tools from cognitive domains and processes is needed to confirm this.

There was also some evidence that elements of executive functioning, such as inhibition, were predictive of social relationships, whereas general visual processes, including visuospatial, visuomotor and attention, were predictive of ADLs. However, as assessment in all domains reviewed were found to be predictive of multiple functional outcomes, these differential relationships are unclear. Generally, assessments tapping into visual skills and executive function (though not language initiation) are likely to be clinically useful for predicting functional outcomes post-stroke. I recommend future studies to investigate the

relationships between multiple domain-specific cognitive assessment tools and specific functional outcomes post-stroke. The selection of assessment tools should consider domains and subdomains involved and effect sizes should be reported to allow relationships to be compared. Preferably the same study should look at multiple cognitive domains/subdomains and multiple functional outcomes, as this allows sample, assessment time-points and tools to remain constant. Future research should be population-based and control for potential confounds such as age and stroke severity. Additional investigation into the impact of assessment time-point on the predictive validity of results would also be clinically useful.

5 References

Aaronson, L. S. (1989). A cautionary note on the use of stepwise regression. *Nursing Research, 38*(5), 309-311.

*Adamit, T., Maeir, A., Assayag, E. et al. (2015). Impact of first-ever mild stroke on participation at 3 and 6 month post-event: The TABASCO study. *Disability and Rehabilitation: An International, Multidisciplinary Journal, 37*(8), 667-673.

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.

Ballard, C., Rowan, E., Stephens, S. et al. (2003). Prospective follow-up study between 3 and 15 months after stroke: improvements and decline in cognitive function among dementia-free stroke survivors> 75 years of age. *Stroke, 34*(10), 2440-2444.

Barker-Collo, S., & Feigin, V. (2006). The impact of neuropsychological deficits on functional stroke outcomes. *Neuropsychology Review, 16*(2), 53-64.

*Bertolin, M., Van Patten, R., Greif, T., & Fucetola, R. (2018). Predicting cognitive functioning, activities of daily living, and participation 6 months after mild to moderate stroke. *Archives of Clinical Neuropsychology, 33*(5), 562-576.

Blomgren, C., Samuelsson, H., Blomstrand, C. et al. (2019). Long-term performance of instrumental activities of daily living in young and middle-aged stroke survivors—Impact of cognitive dysfunction, emotional problems and fatigue. *PloS one,* 14(5), e0216822.

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science,* 25(1), 7–29.

*Cumming, T. B., Brodtmann, A., Darby, D., & Bernhardt, J. (2014). The importance of cognition to quality of life after stroke. *Journal of Psychosomatic Research, 77*(5), 374-379.

*Desrosiers, J., Demers, L., Robichaud, L. et al. (2008). Short-term changes in and predictors of participation of older adults after stroke following acute care or rehabilitation. *Neurorehabilitation and Neural Repair, 22*(3), 288-297.

Di Monaco, M., Schintu, S., Dotta, M. et al. (2011). Severity of unilateral spatial neglect is an independent predictor of functional outcome after acute inpatient rehabilitation in individuals with right hemispheric stroke. *Archives of Physical Medicine and Rehabilitation, 92*(8), 1250-1256.

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of epidemiology and community health, 52*, 377-384.

Duncan, P. W., Lai, S. M., Bode, R. K. et al. (2003). Stroke Impact Scale-16: A brief assessment of physical function. *Neurology,* 60(2), 291-296.

Ferro, M. A., & Speechley, K. N. (2009). Depressive symptoms among mothers of children with epilepsy: a review of prevalence, associated factors, and impact on children. *Epilepsia, 50*(11), 2344-2354.

*Hochstenbach, J. B., Anderson, P. G., van Limbeek, J., & Mulder, T. T. (2001). Is there a relation between neuropsychologic variables and quality of life after stroke? *Archives of Physical Medicine and Rehabilitation, 82*(10), 1360-1366.

*Kapoor, A., Lanctot, K. L., Bayley, M. et al. (2019). Screening for Post-Stroke Depression and Cognitive Impairment at Baseline Predicts Long-Term Patient-Centered Outcomes After Stroke. *Journal of Geriatric Psychiatry and Neurology, 32*(1), 40-48.

Kennedy, J. E., Clement, P. F., & Curtiss, G. (2003). WAIS-III processing speed index scores after TBI: the influence of working memory, psychomotor speed and perceptual processing. *The Clinical Neuropsychologist,* 17(3), 303-307.

Kinsella, E. L., Grace, J. J., Muldoon, O. T., & Fortune, D. G. (2015). Post-traumatic growth following acquired brain injury: a systematic review and meta-analysis. *Frontiers in Psychology, 6*, 1162.

*Laakso, H. M., Hietanen, M., Melkas, S. et al. (2019). Executive function subdomains are associated with post-stroke functional outcome and permanent institutionalization. *European Journal of Neurology, 26*(3), 546-552.

*Larson, E. B., Kirschner, K., Bode, R. K. et al. (2003). Brief Cognitive Assessment and Prediction of Functional Outcome in Stroke. *Topics in Stroke Rehabilitation Winter, 9*(4), 10-21.

Mahoney, F. I., & Barthel, D. W. (1965). Functional evaluation: the Barthel Index: a simple index of independence useful in scoring improvement in the rehabilitation of the chronically ill. *Maryland state medical journal.*

Middleton, L. E., Lam, B., Fahmi, H. et al. (2014). Frequency of domain-specific cognitive impairment in sub-acute and chronic stroke. *NeuroRehabilitation, 34*(2), 305-312.

Mole, J. A., & Demeyere, N. (2020). The relationship between early post-stroke cognition and longer term activities and participation: a systematic review. *Neuropsychological Rehabilitation, 30*(2), 346-370.

*Narasimhalu, K., Ang, S., De Silva, D. A. et al. (2011). The prognostic effects of poststroke cognitive impairment no dementia and domain-specific cognitive impairments in nondisabled ischemic stroke patients. *Stroke, 42*(4), 883-888.

Nijboer, T., Van de Port, I., Schepers, V. et al. (2013). Predicting functional outcome after stroke: the influence of neglect on basic activities in daily living. *Frontiers in Human Neuroscience, 7*, 182.

*Nys, G., Van Zandvoort, M., De Kort, P. et al. (2005). The prognostic value of domain-specific cognitive abilities in acute first-ever stroke. *Neurology, 64*(5), 821-827.

*Nys, G. M. S., van Zandvoort, M. J. E., van der Worp, H. B. et al. (2006). Early cognitive impairment predicts long-term depressive symptoms and quality of life after stroke. *Journal of the Neurological Sciences, 247*(2), 149-156.

*Ownsworth, T., & Shum, D. (2008). Relationship between executive functions and productivity outcomes following stroke. *Disability and Rehabilitation: An International, Multidisciplinary Journal, 30*(7), 531-540.

*Park, J., Lee, G., Lee, S. U., & Jung, S. H. (2016). The impact of acute phase domain-specific cognitive function on post-stroke functional recovery. *Annals of Rehabilitation Medicine, 40*(2), 214-222.

*Park, S. H., Sohn, M. K., Jee, S., & Yang, S. S. (2017). The characteristics of cognitive impairment and their effects on functional outcome after inpatient rehabilitation in subacute stroke patients. *Annals of Rehabilitation Medicine, 41*(5), 734-742.

Rankin, J. (1957). Cerebral vascular accidents in patients over the age of 60: II. Prognosis. *Scottish medical journal*, 2(5), 200-215.

Reitan, R. M. (1955). The relation of the Trail Making Test to organic brain damage. *Journal of Consulting Psychology*, 19(5), 393–394.

Rossini, E. D., & Karl, M. A. (1994). The Trail Making Test A and B: a technical note on structural nonequivalence. *Perceptual and motor skills,* 78(2), 625-626.

Sánchez-Cubillo, I. 1., Periáñez, J. A., Adrover-Roig, D. et al. (2009). Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society: JINS,* 15(3), 438.

*Shea-Shumsky, N. B., Schoeneberger, S., & Grigsby, J. (2019). Executive functioning as a predictor of stroke rehabilitation outcomes. *The Clinical Neuropsychologist, 33*(5), 854-872.

Teasell, R. M. D., & Hussein, N. (2016). Clinical consequences of stroke. Evidence-Based Review of Stroke Rehabilitation. *Ontario: Heart and Stroke Foundation and Canadian Stroke Network*, 1-30.

Uanhoro, J. O., Wang, Y., & O'Connell, A. A. (2019). Problems With Using Odds Ratios as Effect Sizes in Binary Logistic Regression and Alternative Approaches. *The Journal of Experimental Education*, 1-20.

*Van Zandvoort, M. J. E., Kessels, R. P. C., Nys, G. M. S. et al. (2005). Early neuropsychological evaluation in patients with ischaemic stroke provides valid information. *Clinical Neurology and Neurosurgery, 107*(5), 385-392.

Wade, D. T., Wood, V. A., & Hewer, R. L. (1985). Recovery after stroke--the first 3 months. *Journal of Neurology, Neurosurgery & Psychiatry, 48*(1), 7-13.

*Wagle, J., Farner, L., Flekkøy, K. et al (2011). Early post-stroke cognition in stroke rehabilitation patients predicts functional outcome at 13 months. *Dementia and Geriatric Cognitive Disorders, 31*(5), 379-387.

Part Two: Empirical Report

Investigating the validity of the Short Parallel Assessment of Neuropsychological Status (SPANS) as a diagnostic Screen in an acquired brain injury sample:

Assessing the impact of time-since-injury and level of education on classification statistics

Investigating the validity of the Short Parallel Assessment of Neuropsychological Status (SPANS) as a diagnostic Screen in an acquired brain injury sample: Assessing the impact of time-since-injury and level of education on classification statistics

Abstract

Background/Objectives

The SPANS is a global and domain-level neuropsychological assessment tool designed for use in acute acquired brain injury (ABI) settings. The current study assessed the ability of the SPANS to discriminate between participants who had suffered an ABI and an age and education matched healthy norm sample. This ability was assessed in an acute (≤ 6 months post-ABI) and long-term (≥ 1-year post-ABI) sample to assess the validity of the SPANS as a global and domain-specific cognitive screen at these timepoints, and the impact of time-since-injury on cut-off scores and levels of sensitivity and specificity. The impact of level of education on classification statistics was also assessed and the validity of the selection of unitary cut-off scores discussed.

Method

Receiver Operating Characteristic (ROC) curves were plotted to assess the ability of the SPANS total score and individual index scores to discriminate between healthy and brain-injured participants using the following clinical samples: ABI ≤ 6 months, ABI ≥ 1 year, high school educated, college/vocationally educated, university educated. Classification statistics were plotted for each ROC curve and optimal cut-off scores chosen and compared.

Results

The SPANS total score was an outstanding discriminator for the acute ABI sample and an excellent discriminator for the long-term sample. Individual SPANS indices were all excellent-outstanding discriminators in the acute sample and were poor-excellent in the long-term sample. High levels of sensitivity and specificity were achievable for cut-off scores for the SPANS total score in all samples, though these were lower for individual index scores in the long-term population. The memory index was the best discriminator in the long-term sample and orientation and conceptual flexibility indices the worst. Education level and time-since injury significantly impacted optimal cut-off score values.

Conclusions

Results indicate that the SPANS is a valid global and domain-specific cognitive screen in acute ABI samples, and a valid global screen for longer-term ABI samples. Its validity as a domain-specific screen in long-term ABI is inconclusive. The clinical sample used did not have confirmed cognitive impairments and functioning in many cognitive domains may have recovered at this time post-injury. Results indicate that impairments in memory endure long-term post-ABI and should be a focus of rehabilitation planning. Impairments in all other domains were still present and detectable in the long-term sample, indicating many areas of cognitive impairment endure in some patients. Optimal cut-off scores are influenced by patient factors and clinical need, therefore, reliance on unitary cut-off scores could reduce the validity of clinical screens and result in the misdiagnosis of impairments.

1 Introduction

1.1 Cognitive assessment after brain injury

Acquired brain injury (ABI) is a term used to describe many types of insult that result in damage to the brain, including internal injuries caused by events such as haemorrhage or anoxia, and external force injuries such as those caused by falls or road traffic accidents, referred to as traumatic brain injuries (TBI; Greenwald et al., 2003). Worldwide, ABI causes long-term disability in 150-200 people per million annually, commonly resulting in physical, social and/or cognitive deficits (Fleminger & Ponsford, 2005). Generalised cognitive impairment post-injury has been demonstrated to be one of the most important predictors of long-term rehabilitation and functional outcomes (e.g. Narasimhalu et al., 2009; Zinn et al., 2004). This highlights the need for accurate assessment of a person's cognition post-injury.

Cognitive impairment does not occur in a unitary 'domain' after ABI however, but occurs in different patterns across many domains of cognitive functioning, depending on the location and type of damage (McCauley et al., 2014). These domains are commonly referred to, or defined as, language, memory, attention, visuospatial perception and executive function, among others. Domain-specific cognitive impairments have been found to be particularly important for different functional outcomes. For example, visuo-perceptual and verbal reasoning impairments were found to be better predictors of functional outcomes than other domains (Sigurdardottir et al., 2009).

A useful cognitive assessment for ABI populations would therefore require the ability to accurately detect and classify cognitive impairment at a domain-specific and global level. This information is necessary to aid in the creation of targeted rehabilitation plans (e.g. managing poor memory) which can ameliorate some of the impact of cognitive impairment

(Weicker et al., 2016). It is also needed to inform safe discharge decisions, and the provision of care and allocation of resources.

Neuropsychological tests for use in brain-injured populations are, by necessity, brief and simple (Nabors et al., 1997). Many of the currently used cognitive assessment tools are adapted dementia screening tools which deliver a single 'global' cognitive domain score with an advised 'cut-off ', which applies to all examinees (Cartoni & Lincoln, 2005; Folstein et al., 1975; Mathuranath et al., 2006; Nasreddine et al., 2005). Research has generally, albeit inconsistently, demonstrated that these tests are reliable and valid for this purpose; some have shown predictive validity and relationship to functional outcomes in ABI populations (e.g. Cartoni & Lincoln, 2005; Chiti & Pantoni, 2014). The nature of the 'pass or fail' paradigm, however, risks being under-informative in regard to distinct cognitive domains. Further testing would be required in order to provide information on the pattern and severity of domain-specific impairments that is required for targeted rehabilitation plans.

The Short Parallel Assessment of Neuropsychological Status (SPANS; Burgess, 2014) is a short battery neuropsychological assessment tool that was created to address these shortcomings and be a stand-alone, 'sufficient' screen, or brief assessment, in acute ABI populations. Its design aim was to provide normed scores across multiple cognitive domains, providing sufficient information to inform rehabilitation plans, whilst not placing an excessive test-taking burden on patients. The SPANS design was informed by common referral questions received by neuropsychologists in acute ABI neurorehabilitation settings. Its intended use is in these acute settings, whilst also having the potential for utility in longer-term, neurological outpatient settings. There are two alternate versions making it possible to re-assess patients at a later time-point to provide information on cognitive recovery and rehabilitation progress.

The SPANS provides standardised and co-normed scores across seven indices chosen based on neuropsychological theory related to ABI. These are: The orientation index (ORI), the memory/learning index (MLI), the language index (LAI), the visuo-motor performance index (VPI), the attention/concentration index (ACI), the efficiency index (EFI; assesses psychomotor speed, speed of making mental calculations, visual scanning efficiency, and reaction time) and the conceptual flexibility index (CFI). For detailed information on the contents of each index, see appendix F.

The SPANS indices have shown good internal reliability and convergent validity in several studies carried out to date (Burgess, 2014) indicating that they are reliable and valid measures of the cognitive domains they aim to assess. For a screening or diagnostic test to be useful however, it must also be able to reliably discriminate between people who have the condition in question and those who do not. A previous ANOVA study indicated that the SPANS is likely a valid discriminator of degree of cognitive impairment, with all indices successfully distinguishing between an ABI (less than one year post-injury) and healthy norm sample, and all but one successfully distinguishing between these samples and a sample of patients with long-term neurological conditions, in a 3 x 7 group means comparison (Burgess, 2014).

While this is good initial evidence of discriminative ability, further analysis is needed to provide estimates of sensitivity and specificity, which are important for a diagnostic tool. A useful clinical screening tool must demonstrate adequate sensitivity, keeping the rate of 'false negative' results (i.e. failing to detect the condition when it exists) low. At the same time, it must maintain adequate specificity, keeping the rate of 'false positive' results (i.e., indicating existence of the condition when it does not in fact exist) low. The current study aims to provide estimates of sensitivity and specificity for the SPANS by carrying out a

Receiver Operating Curve (ROC) analysis to provide these estimates. The same sample used for the ANOVA study (Burgess, 2014) will be utilised with the addition of more recent data collected since this analysis, and with the sample of long-term neurological conditions excluded. This decision was taken as the long-term neurological conditions in the SPANS sample were gradual and neurodegenerative in nature, which may influence the ability of the SPANS to detect them, therefore skewing results or any conclusions drawn for its use after ABI.

To provide sensitivity and specificity estimates a particular 'cut-off score' must be derived. Most cognitive screening tools provide one cut-off score to be used for all patients, with current guidance indicating that sensitivity be prioritised over specificity for a clinical diagnostic test, in order to prevent positive cases from being missed (Lincoln et al., 2003). However, there are problems with the aim of selecting a unitary cut-off point. Firstly, the desired balance of sensitivity and specificity is situation specific. There may be situations clinicians face where cut-off scores that favour sensitivity over specificity may be costly or inconvenient, in terms of referring for services where they may not be needed, or the human cost of unnecessary anxiety.

Secondly, patient variables can impact cognitive test performance, meaning the sensitivity and specificity of a cut-off score is unlikely to remain static in different groups. Age and education level are two of the most significant moderating factors of neuropsychological test performance and have been found to account for up to 49% of variance in scores on some assessments (Malek-Ahmadi et al., 2015). Research has also found that higher education levels and lower age are both linked to better cognitive recovery post-ABI (Fortune et al., 2016; Smania et al., 2013). Additionally, single cut-offs are not likely to be optimal for both acute and long-term populations. In the acute phase post-injury

impairments are at their most severe, with an accelerated recovery taking place within the first 5-6 months (Christensen et al., 2008). This means that cut-off scores for long-term populations may need to be lower in order to have equal sensitivity.

1.2 Aims of current study

The aim of the current study was to assess the ability of the SPANS total score and individual index scores to discriminate between brain-injured and healthy participants. This was investigated in both an acute (less than 6 months post-injury) and long-term (over 1-year post-injury) sample and for different levels of education, allowing the impact of these factors on optimal cut-off scores to be assessed. Cut-off scores were selected allowing the prioritisation of either sensitivity or specificity, making them adaptable to different clinical needs. Brain-injured and healthy samples were age and education matched due to the influence these factors have on cognitive assessment scores (Malek-Ahmadi et al., 2015).

A potential weakness of the sample used is that impairment was not confirmed in the clinical sample using 'gold-standard' testing. This could have resulted in lower levels of discriminative ability as impairment was unlikely to be present in all domains, particularly in the long-term sample where recovery would be expected (Christenson et al., 2008). Evidence does show however, that post-ABI cognitive impairment is generally an enduring problem. Vaishnavi et al. (2009) found that some form of cognitive impairment after a moderate-severe ABI was present at 2-years post injury and Sigurdardottir et al. (2020) found impairments persisted, and were relatively stable, at 10 years post-injury. Even mild brain injury has been shown to result in some cognitive impairments that are still detectable at 12-months post injury (Bedard et al., 2020), though many domains recover rapidly after a mild-ABI (Schretlen & Shapiro, 2003). This indicates that some form of impairment was likely to be

present in the long-term ABI sample, making it a valid way to measure the discriminative ability of the SPANS total score.

Different cognitive domains, however, show different patterns of recovery. For example, Christensen et al. (2008) found that functioning recovered more steeply in areas of memory, executive functioning, verbal ability and dexterity, and Shiel et al. (2000) found that orientation and 'simple' memory tend to return soon after stroke. Impairments in attention on the other hand have been found to endure in mild-brain injury samples (Vanderploeg, et al., 2005) and Carroll et al. (2020) found problems with episodic and working memory can endure post 2-years. The discriminative ability of individual SPANS indices in the long-term therefore likely indicated which domain-specific impairments are present and enduring in this sample, rather than be an accurate measure of their screening ability.

1.2.1 Main Aims and hypotheses

1) The ability of the SPANS total score to discriminate between a healthy, age and education-matched, sample and acute and long-term ABI samples was assessed, allowing the impact of time-since-injury to be investigated. It was hypothesised that the SPANS would demonstrate better discriminative ability, higher levels of sensitivity and specificity, and higher optimal cut-off scores in the acute population due to the likely higher levels of impairment present.
2) The ability of individual SPANS indices to discriminate between acute and long-term ABI samples and healthy age and education matched participants was assessed. This helped to indicate the SPANS' validity as a domain-level cognitive screen and inform the pattern of recovery across different cognitive domains post-ABI. It was hypothesised that all indices would be good discriminators in the acute sample, but that the orientation, language, visuomotor and conceptual flexibility indices would be

poor discriminators in the long-term sample, while attention and memory would be the best. This was based on differential patterns of recovery found by previous research (e.g. Christensen et al., 2008; Shiel et al., 2000).

3) The impact of level of education on cut-off scores for the SPANS was assessed. It was hypothesised that cut-off scores would be higher for participants with higher levels of education, due to previous research showing a link between higher levels of education and higher scores on cognitive assessment (Malek-Ahmadi et al., 2015) and further research indicating the role of education level in cognitive recovery post-ABI (Smania et al., 2013).

2 Method

2.1 Participants

Participants were a subset[3] of those recruited by the SPANS author, Dr Gerald Burgess, for use in the validation of the SPANS. A proportion of this sample were used for published analyses in the SPANS manual (Burgess, 2014), though some data has been collected subsequently. Two samples were used for analysis, a clinical sample of participants who had suffered an ABI and a sample of healthy age-matched participants.

Participants in the clinical sample were 104 non-consecutive patients referred to inpatient and outpatient neuropsychology services, across two large cities in the UK, from 2008 to 2019. Data were collected as part of routine care and participants were enrolled only if administration of the SPANS was relevant to the referral and their ongoing care. Data were only collected if the patient was deemed able to complete the full SPANS assessment, this

[3] Participants with long-term neurological conditions were removed and data was age and education matched resulting in the loss of some cases, see data preparation section.

necessarily excluded patients who were in a coma, were physically disabled, blind or significantly aphasic.

As an additional inclusion criterion for the present study, participants were included only if they had suffered an ABI. As a result, a subset of participants with long-term neurological conditions were excluded. Participants were a wide range of age and education levels and were varied in the severity of their injuries with scores ranging from 'mild' to 'severe' on the Glasgow Coma Scale (GCS; Teasdale & Jennett, 1974).

The sample of healthy adults consisted of contacts of the research team including patients' family members, NHS staff and university employees. Inclusion/exclusion criteria for both the clinical and healthy sample were English as a first language and education in a western culture, having been demonstrated to impact on SPANS scores (Tan & Burgess, 2018). An additional inclusion criterion for the healthy sample was no history of a previous neurological condition or ABI. A summary of the demographic and clinical information for both samples is displayed in Table 1.

Table 1: Demographic details of ABI and healthy norm samples

	Whole ABI sample *(N= 104)*	Healthy norm sample *(N = 104)*
Mean age (SD)	43.99 (14.78)	43.82 (15.16)
Age Range	18-84	18-82
Sex M/F	86/18	48/56
Education		
Secondary education	39(37.5%)	39 (37.5%)
College or vocational training	46 (44.2%)	46 (44.2%)
University Educated	19 (18.3.%)	19 (18.3%)
Injury Type		
TBI	55 (53.8%)	N/A

Anoxic/hypoxic	13 (12.5%)	N/A
Haemorrhage	16 (15.4%)	N/A
Stroke	10 (9.6%)	N/A
Other ABI	9 (8.7%)	N/A
Weeks since injury Mean (SD)	92.23 (158.8)	N/A
GCS Mean (SD)	7.65 (3.88)	N/A
Total SPANS score Mean (SD)	194.03 (47.03)	242.51 (10.56)

Pair-wise analyses were run to examine whether there were any significant differences in demographic data between the clinical and control group. A paired t-test showed no significant difference in mean age between the samples however a McNemar's test showed there was a significant difference in gender (see Appendix G for both analyses). Differences in education level were not explored due to this being matched on an exact basis, meaning there could not be any differences between the two samples, as demonstrated in the descriptive statistics. As gender differed significantly between the samples two-way repeated measures ANOVAs were run to see whether this gender difference impacted on scores for the SPANS total and each index. All ANOVA's showed a significant main effect of sample type on scores, but none found a significant effect of sex or a significant interaction between sex and sample type (see Appendix H). This indicates that this difference in gender ratios between the samples is unlikely to have impacted on scores or interfered with results from the ROC analyses.

2.2 Measures

2.2.1 SPANS

A summary of the relevant reliability and validity statistics for the SPANS available to date are displayed in Table 2. These include internal validity statistics, test-retest reliability and convergent validity with correlations of .50 and above.

Table 2: SPANS reliability and validity statistics

Index	ORI	ACI	LAI	MLI	VPI	ECI	CFI
Cronbach's Alpha							
Form A	.80	.82	.84	.89	.85	.83	.70
Form B	.77	.85	.88	.91	.85	.87	.77
Combined	.79	.83	.86	.90	.85	.85	.73
Test Re-test							
Reliability coefficient	.93	.88	.97	.95	.89	.88	.74
1st administration mean (SD)	21.88 (.31)	44.33 (1.43)	51.42 (1.88)	63.08 (2.75)	67.17 (1.90)	46.92 (1.17)	27.58 (.52)
2nd administration mean (SD)	21.96 (.14)	45.25 (1.22)	51.75 (1.06)	63.50 (2.48)	67.75 (1.71)	47.33 (.99)	27.58 (.52)
Convergent Validity							
WAIS Verbal IQ	N/A	N/A	.657	N/A	N/A	N/A	N/A
WAIS Verbal comprehension	N/A	N/A	N/A	N/A	N/A	N/A	N/A
WAIS Working Memory	N/A	.626	.509	N/A	N/A	N/A	N/A
WAIS Performance IQ	N/A	N/A	N/A	N/A	.693	.673	.513
WAIS Perceptual Organisation	N/A	N/A	N/A	N/A	.674	.595	.605
WAIS Processing Speed	N/A	N/A	N/A	N/A	.567	.581	N/A
WMS Auditory Immediate	N/A	N/A	N/A	.563	N/A	N/A	N/A
WMS Auditory Delayed	N/A	N/A	N/A	.712	N/A	N/A	N/A
Rey Complex Figure Immediate	N/A	N/A	N/A	.667	.684	.637	N/A
Rey Complex Figure Delayed	N/A	N/A	N/A	.652	.689	.599	N/A
Trail Making Test A	N/A	-.564	N/A	-.555	-.574	-.612	N/A
Trail Making Test B	-.785	-.753	N/A	N/A	-.901	-.857	N/A

Internal consistency statistics ranged from satisfactory to excellent suggesting that SPANS indices are a trustworthy measure of a single cognitive construct. Correlation coefficients also ranged from adequate to excellent suggesting SPANS index scores are trustworthy across multiple administrations using both versions. Convergent validity results suggest that SPANS indices correlate most highly with measures (chosen for their predictive validity for functional outcomes) purporting to measure the same construct and negatively with those measures that are theoretically least similar.

2.3 Design and procedure

The study was a retrospective cohort analysis using archival data for use in the validation of the SPANS. Data were collected during routine clinical practice. The SPANS was always the first psychometric test administered to patients to avoid any practice effects from other cognitive assessments and efforts were made to avoid interruptions or pauses. Standardised testing procedure was used as described in the test manual (Burgess, 2014).

A subset of 27 participants from the clinical sample were re-tested with an alternate version of the SPANS, with SPANS A form administered in the acute phase and SPANS B at a later time-point, as would be done in normal clinical practise. The testing procedure was the same for the healthy norm sample, but SPANS form was randomly assigned. A subset of 53 participants from the healthy norm sample were also re-tested with the alternate form. All re-test data (both clinical and norm) were removed from the dataset before analysis, so that the same participant was not included twice as separate data points.

2.4 Ethical considerations

Full ethical approval was given by the NHS Research Ethics Committee (REC) for the data to be collected and for anonymised data to be used in any secondary analysis approved by the lead researcher, Dr Burgess (see Appendix I for REC approval letter).

2.5 Data analysis

2.5.1 Data preparation

Anonymised data were provided to the author of the current study on an IBM SPSS database. Data were thoroughly screened by the author with data not meeting the inclusion/exclusion criteria (not ABI),having missing data needed for the main analyses, or being a re-test of the same participant, removed. Where re-test data were removed, time-point 1 data were kept. Participants from the clinical sample were then pair-matched with participants from the norm sample, based on age and education, using the SPSS matching function. Education level was set as an exact match, with participants coded with numbers (1 = secondary education; 2 = some college or vocational training; 3 = university degree). Age was set with a 5-year variance, this number was reached using an iterative process in order to identify the highest number of matches whilst maintaining no significant variance in age between the two samples. This was assessed using a *t*-test means comparison to check for significant difference in age between the two samples. The final case matched sample consisted of 104 clinical cases and 104 matched healthy norms, see Figure 1 below for a summary of the data preparation process.

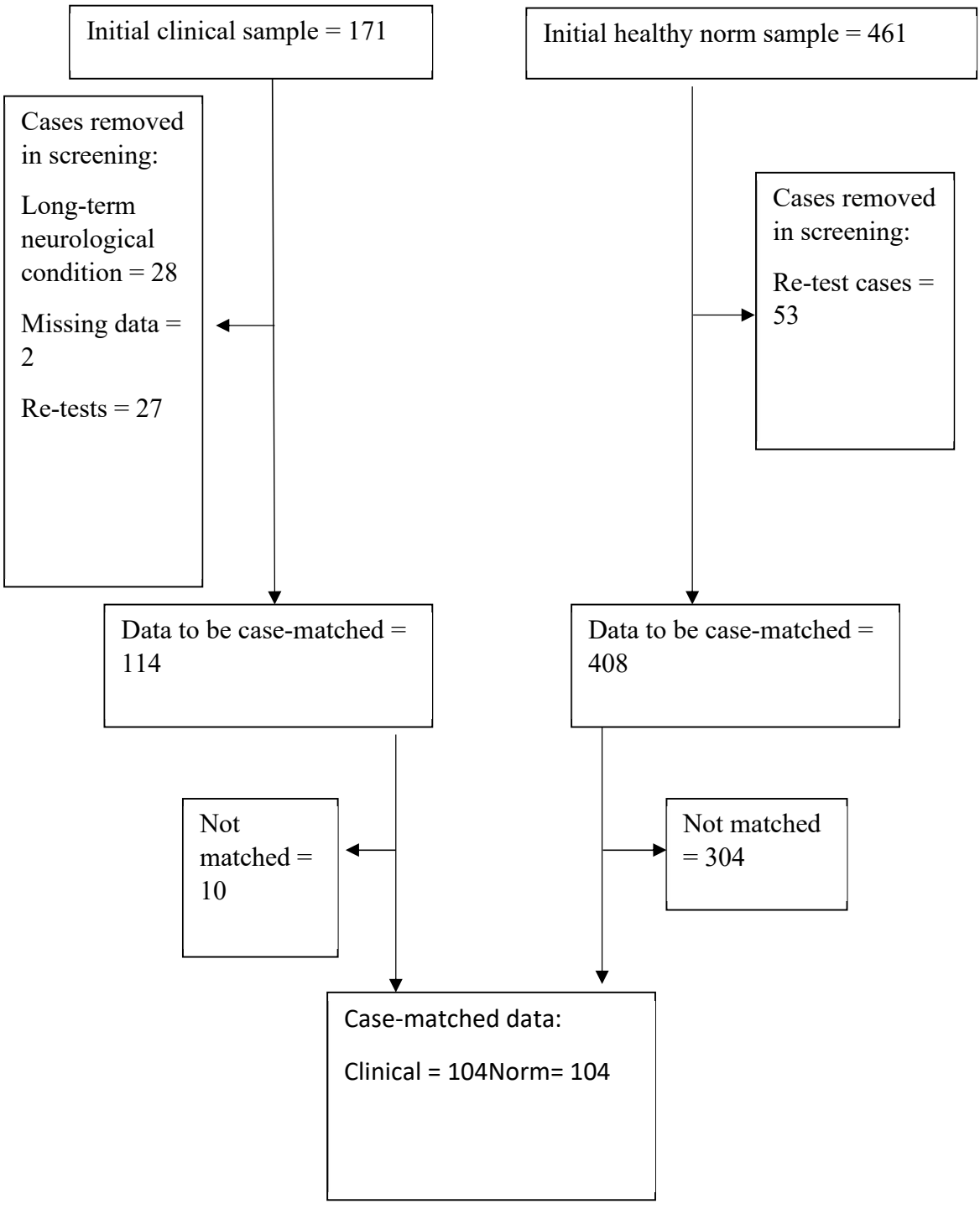

Figure 1: Flowchart of the data preparation process

2.5.2 Power analysis

A power analysis was carried out using Medcalc with the area under the ROC curve estimated as .80, due to the SPANS being expected to be able to effectively discriminate participants with a brain injury from healthy controls, based on the previous ANOVA study results (Burgess, 2014). The null hypothesis value was set at 0.5 (chance level), the alpha at .05 and with a β-level of 0.20 (80% power). The power analysis indicated a total sample size of 26 would be required in order to reject the null hypothesis in a 1:1 sample.

2.5.3 Receiver Operating Characteristic (ROC) curves

ROC curves were plotted for the SPANS total score and its seven index scores, using SPSS, to examine their ability to successfully discriminate between clinical ABI samples (≤ 6-months post injury, ≥1 -year post injury and three levels of education) and an age and education matched healthy sample. ROC curves plot sensitivity (true positive rate) against false positive rates (1-specificity), with the left-hand corner of the axis representing 'perfect' discrimination and the diagonal line across the centre of the graph indicating chance level discrimination. ROC curve analysis also produces an Area Under Curve (AUC) statistic, which shows the overall discriminative validity of a diagnostic test, with a score of 1 indicating perfect discrimination. Hosmer and Lemeshow (2000) posit that an AUC of ≥ .90 be considered outstanding, ≥.80 excellent and ≥.70 acceptable.

2.5.4 Classification statistics and selection of optimal cut-off scores

Classification statistics were produced as part of the ROC analysis to provide estimates of sensitivity and specificity for different cut-off scores. Sensitivity refers to the true positive rate, in other words what proportion of the people who are impaired would be

correctly identified as impaired by this test score. Specificity refers to the true negative rate, or, how many people are correctly classified as unimpaired by this test score. Youden's J (Youden, 1950), positive and negative likelihood ratios (+/- LR) and positive and negative predictive values (PPV/NPV) were then calculated using SPSS and Microsoft Excel, to aid in the selection of optimal cut-off points.

2.5.5 Youden's J Index

Youden's J index (max: sensitivity + specificity – 1; Youden, 1950) provides a numerical value indicating the overall number of false positive and false negative results that would be produced by a particular cut-off, in a given sample. A Youden's value of 1 indicates perfect discrimination with no false positives or false negatives. Youden's J equation values false positives and false negatives equally therefore the cut-off score with the highest Youden's value indicates the best trade-off between sensitivity and specificity, with the fewest overall number of false negatives and positives combined. Lincoln et al. (2003) suggested that an acceptable diagnostic test should demonstrate a sensitivity of ≥ 80% and specificity of ≥ 60%. Cut-offs were therefore chosen which had the highest Youden's index but also fell within Lincoln et al.'s (2003) criteria, where possible. A second cut-off point was also chosen which prioritised specificity, as discussed in the introduction, while maintaining a high Youden's score and not over-compromising on sensitivity.

2.5.6 Likelihood ratios and predictive values

Likelihood ratios and positive and negative predictive values were calculated as a descriptive indicator of the overall utility of chosen cut-offs. PPVs show the percentage of positive test results that would be true positives, while NPVs show the percentage of 'true'

negative results. Positive and negative likelihood ratios indicate how much more likely a person with the condition (in this case an ABI) is to have a positive or negative test result.

2.5.7 DeLong et al.'s (1988) method for comparing ROC curves

ROC curve AUC's were compared using DeLong et al.'s (1988) statistical analysis on MedCalc to check for significant differences in AUC statistics. The analysis for comparing related ROC curves was used to assess differences in AUC's for different indices in the same sample (e.g. acute) and the analysis for comparing independent ROC curves was used to examine differences between ROC curves using different samples (e.g. acute vs long-term).

3 Results

3.1 Acute vs long-term ABI sample

A total of 130 participants (65 clinical, 65norm) were included in the acute sample and a total of 62 participants (31 clinical, 31 norm) were included in the long-term sample. Table 3 displays the descriptive statistics for the two clinical samples.

Table 3: Descriptive statistics of the acute and long-term ABI sample

	≤ 6-months post-injury *(N = 65)*	≥1-year post-injury *(N = 31)*
Age Mean (SD)	42.88 (15.40)	44.54 (14.74)
Age Range	18-74	2082
Education		
Secondary education	25 (38.4%)	10 (32.3%)
College or vocational training	28 (43.1%)	16 (51.6%)
University Educated	12 (18.5%)	5 (16.1%)
Injury Type		
TBI	33 (50.8%)	20 (64.5%)
Anoxic/hypoxic	11 (16.9 %)	1 (3.2%)
Haemorrhage	12 (18.5%)	3 (9.7%)
Stroke	7 (10.8%)	2 (6.5%)
Other ABI	2 (3.1%)	5 (16.1%)
Weeks since injury Mean (SD)	8.49 (4.39)	281.52 (183.20)
GCS Mean (SD)	7.38 (3.48)	8.71(4.99)
Total SPANS score Mean (SD)	184.66 (48.52)	213.65 (33.63)

There were no particular differences between the two samples for age or education. As would be expected the SPANS total score and GCS were significantly higher in the long-term sample indicating cognitive recovery.

3.1.1 Acute sample (pre-6 Months)

ROC curves for the acute sample are displayed in Figures 2 and 3, with AUC statistics displayed in Table 4.

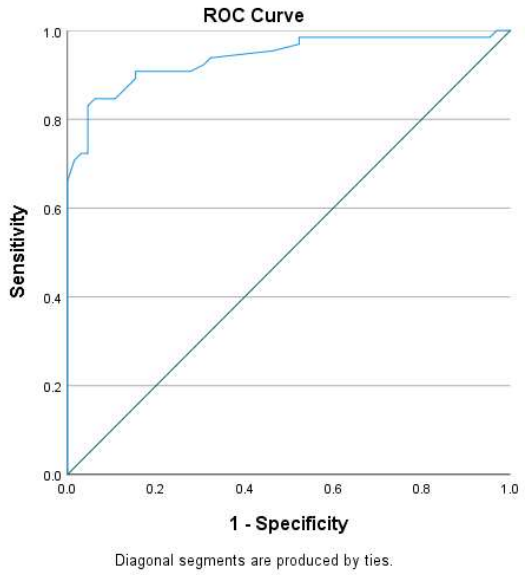

Figure 2: ROC curve for SPANS total score in the acute ABI sample

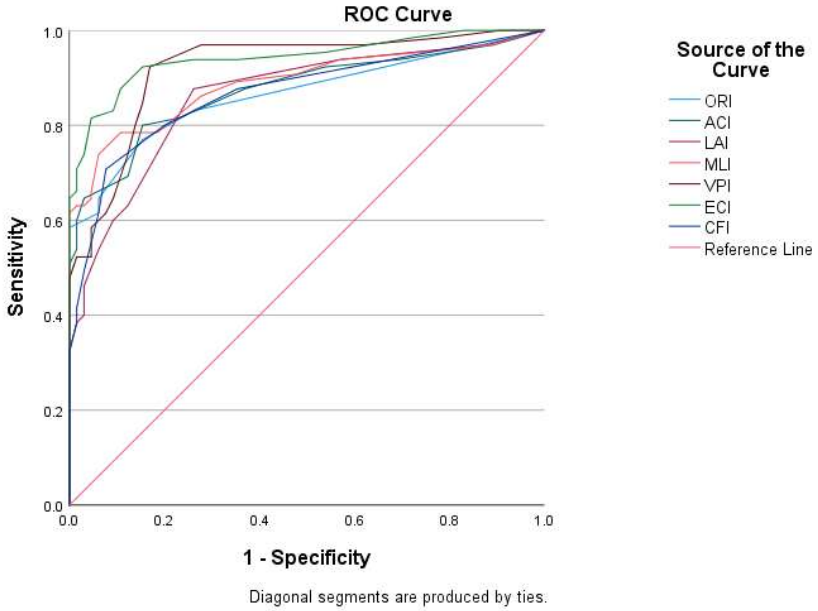

Figure 3: ROC curves for SPANS indices in the acute ABI sample

Table 4: AUC statistics for SPANS indices and total score in acute ABI sample

				95% Confidence Interval	
Test Variable	**AUC**	**Std. Error**	**Significance**	**Lower Bound**	**Upper Bound**
SPANS total	.939	.022	.000	.897	.981
ORI	.864	.033	.000	.799	.930
ACI	.874	.032	.000	.811	.937
LAI	.862	.033	.000	.797	.929
MLI	.889	.030	.000	.830	.948
VPI	,921	.024	.000	.874	.968
ECI	.942	.021	.000	.901	.983
CFI	.868	.033	.000	.804	.932

As can be seen in Figure 2 the curve representing the SPANS total scores is above the diagonal 'line of no interest' which represents the null hypothesis. The curve very closely follows the top left corner of the axis indicating extremely good discrimination. This is supported by an 'outstanding' AUC of .939 which is highly significant $p < .001$. Figure 3 shows ROC curves for each individual SPANS index, all indices have curves above the line of no interest and towards the left corner of the axis indicating good discrimination. This is supported by AUC's ranging from excellent to outstanding all of which were highly significant ($p < .001$). The ECI and VPI had the highest AUCs, indicating better discriminative ability, with the ECI exceeding the SPANS total score with an outstanding AUC of .942. De long et al.'s (1988) statistical test to compare related ROC curves was undertaken on MedCalc to check whether differences between AUCS were significant. The differences in the AUCs for the VPI and ECI and other indices were found to be significantly different, with *p* values ranging from .0003 to .008. Differences between AUCs for the other

indices were non-significant indicating similar levels of discriminative ability. Chosen cut-off scores, Youden's index scores, likelihood ratios and P/NPV are displayed in Table 5.

Table 5: Classification statistics for the SPANS total score and individual indices in acute ABI sample

Index	Prioritised	Cut-off	Sensit-ivity	Specif-icity	Youden	+LR	-LR	PPV	NPV
SPANS Total*	Sensitivity	237	90.8%	80.8%	.715	4.720	0.114	84%	89%
	Specificity	230	84.6%	93.6%	.782	13.200	0.164	93%	86%
ORI*	Sensitivity	21	83.1%	71.8%	.549	2.940	0.235	88%	72%
	Specificity	20	76.9%	83.3%	.603	4.615	0.276	100%	71%
ACI*	Sensitivity	43	87.7%	62.8%	.505	2.358	0.195	70%	84%
	Specificity	40	69.2%	87.2%	.564	5.400	0.352	85%	74%
LAI*	Sensitivity	51	87.7%	73.1%	.608	3.257	0.168	77%	86%
	Specificity	50	63.1%	85.9%	.490	4.472	0.429	82%	70%
MLI*	Sensitivity	61	86.2%	73.1%	.592	16.400	0.384	77%	84%
	Specificity	58	73.8%	92.3%	.662	9.600	0.283	91%	78%
VPI*	Sensitivity	64	92.3%	82.1%	.744	5.142	0.093	87%	92%
	Specificity	61	73.8%	87.2%	.610	5.760	0.3	87%	77%
ECI*	Sensitivity	42	90.8%	82.1%	.728	5.057	0.112	84%	90%
	Specificity	39	81.5%	93.6%	.751	12.720	0.197	95%	84%
CFI*	Sensitivity	27	87.7%	61.5%	,492	2.280	0.2	89%	78%
	Specificity	26	70.8%	91%	.618	7.885	0.321	88%	76%

*Maximum total scores: SPANS Total = 334, ORI = 22, ACI = 46, LAI = 53, MLI = 67, VPI = 70, ECI = 48, CFI = 28

As can be seen in Table 5 the SPANS total score demonstrated high levels of both sensitivity and specificity. Both cut-offs chosen to prioritise sensitivity and specificity show high levels (over 80%) of both, surpassing Lincoln et al.'s (2003) criteria for an acceptable diagnostic test. This is supported by high rates of PPV and NPV. As expected from the ROC curves and AUC of the indices VPI and ECI showed the highest levels of overall sensitivity and specificity, with the ECI in particular showing equivalent levels to the SPANS total

score. score. All indices were able to provide cut-offs meeting Lincoln et al.'s (2003) criteria for an acceptable diagnostic test indicating that they all have good diagnostic ability.

3.1.2 Long-term sample (post 1-year)

ROC curves for the long-term sample can be seen below in Figures 4 and 5 with information on the AUC's and significance level displayed in Table 6.

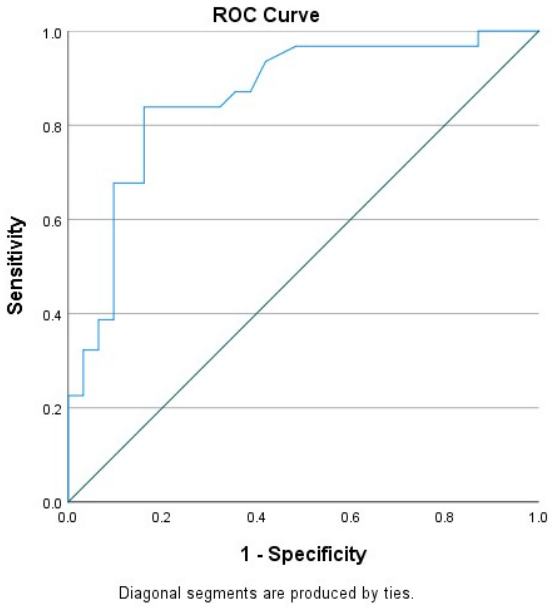

Figure 4: ROC curve for SPANS total score in the long-term ABI sample

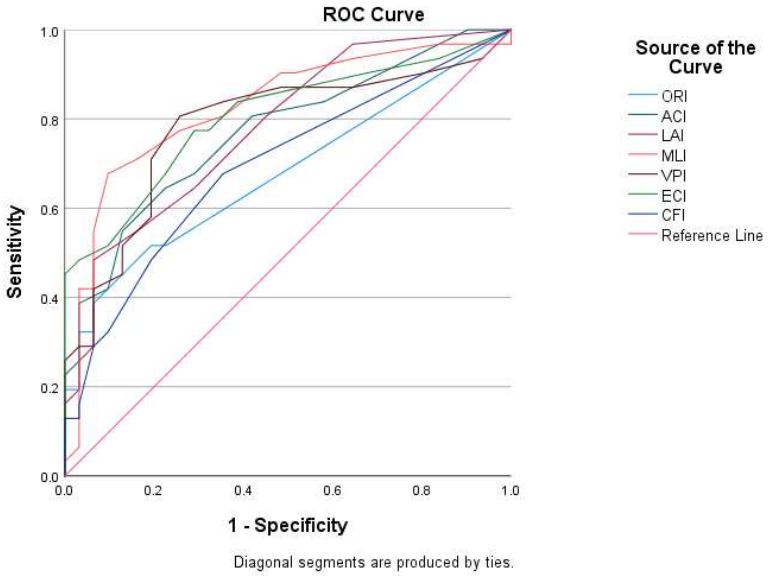

Figure 5: ROC curves for SPANS indices in the long-term ABI sample

Table 6: AUC Statistics for SPANS indices and total score in long-term ABI sample

				95% Confidence Interval	
Test Variable	AUC	Std. Error	Significance	Lower Bound	Upper Bound
SPANS total	.859	.049	.000	.763	.955
ORI	.678	.069	.016	.543	.813
ACI	.772	.060	.000	.655	.888
LAI	.779	.058	.000	.666	.893
MLI	.826	.055	.000	.718	.934
VPI	.785	.061	.000	.665	.905
ECI	.805	.057	.000	.695	.916
CFI	.695	.067	.008	.563	.827

As can be seen in Figure 4 the ROC curve plotted for the SPANS total score in the post 1-year analysis is again above the 'line of no interest' and towards the left hand 'perfect' corner of the axis, indicating a good level of discriminative ability. This is supported by an excellent AUC of .859 with a high significance level ($p < .001$). As expected, this is lower than the outstanding AUC for the acute, injury-prior-to-6-month. sample however, Delong et al.'s (1988) analysis for comparing independent ROC curves indicated that this difference was not statistically significant. The ROC curves for the individual indices are also above the line of no interest, although some of them, particularly the ORI and CFI, are further from the left hand 'perfect' corner indicating poorer discriminative validity. This is further demonstrated by mixed AUC's ranging from 'poor' to 'excellent'. Both the ORI and CFI had 'poor' AUC's while all other indices demonstrated 'acceptable' levels of discriminative validity for this sample except for the MLI and ECI. The MLI and ECI were classified as 'excellent' discriminators, with the MRI having a highly significant AUC of .826, nearly equalling the AUC of the SPANS total score. However, DeLong et al.'s (1988) statistical test to compare related ROC curves demonstrated that the ORI and CFI AUCs were significantly

different only from those of the MRI and ECI indices (p = .0171; p = .0175) as was the case with the other 3 indices (p = .0214;p = .0013; p = 0.351),. This indicates that the MRI and ECI were significantly better discriminators than the other indices in this sample. Chosen cut-offs, sensitivity/specificity, Youden's J, +/-LR and P/NPV's are displayed in Table 7.

Table 7: Classification statistics for the SPANS total score and individual indices in long-term ABI s

Index	Prioritised	Cut-off	Sensit-ivity	Specif-icity	Youden's	+LR	-LR	PPV	NPV
SPANS Total*	Sensitivity	246	93.5%	61.8%	.553	2.446	0.104	71%	90%
	Specificity	230	67.7%	91.2%	.589	7.677	0.353	88%	74%
ORI*	N/A	21	51.6%	82.4%	.340	2.924	0.587	81%	61%
ACI*	Sensitivity	44	80.6%	61.8%	.424	2.109	0.313	68%	76%
	Specificity	42	64.5%	79.4%	.439	3.133	0.446	74%	69%
LAI*	Sensitivity	52	80.6%	55.9%	.365	1.827	0.346	64%	74%
	Specificity	50	54.8%	85.3%	.401	3.729	0.529	81%	66%
MLI*	Sensitivity	61	80.6%	67.6%	.483	2.492	0.286	75%	77%
	Specificity	58	67.7%	91.2%	.589	7.677	0.353	89%	67%
VPI*	Sensitivity	65	80.6%	76.5%	.571	3.427	0.253	78%	80%
	Specificity	64	71%	82.4%	.533	4.021	0.352	81%	74%
ECI*	Sensitivity	45	83.9%	64.7%	.486	2.376	0.249	73%	76%
	Specificity	42	67.7%	79.4%	.472	3.290	0.406	78%	71%
CFI*	N/A	27	67.7%	64.7%	.324	1.919	0.498	68%	60%

*Maximum total scores: SPANS Total = 334, ORI = 22, ACI = 46, LAI = 53, MLI = 67, VPI = 70, ECI = 48, CFI = 28

The SPANS Total score was able to provide a cut-off with a high level of sensitivity and acceptable level of specificity, meeting the criteria of Lincoln et al. (2003) and producing

an extremely good NPV of 90%. By sacrificing some sensitivity, a very high level of specificity was also able to be achieved as an alternative cut-off. As expected, the optimal cut-off for the SPANS total score was 9 points higher than in the acute sample. The majority of individual index scores showed much lower levels of sensitivity/specificity with the ORI and CFI only being able to provide one cut-off score, which maximised sensitivity. The ORI was particularly poor with sensitivity reaching only 51.6%. The MLI, VCI, ECI and ACI however was able to produce cut-offs with an excellent level of sensitivity at over and acceptable level of specificity, as per Lincoln et al.'s (2003) criteria. For the MLI, a second cut-off was able to be produced which gave an excellent specificity of over 90% whilst maintaining a sensitivity of close to 70%. The LAI was not able to meet Lincoln et al.'s (2003) criteria, of a sensitivity of over 80% and specificity of over 60%. Overall PPV and NPV values were lower than for individual indices at 6 months.

3.2 Discriminative ability in different levels of education

The overall sample size for participants with high school level education was 78 (39 clinical, 39 norm) for college/vocational education was 92 (46clinical, 46 norm) and for university educated was 38 (19 clinical, 19 norm). Descriptive statistics for each sample are displayed in Table 8.

Table 8: Descriptive statistics of clinical samples for three levels of education

	High school Educated	Some college/vocational education	University educated
Age Mean (SD)	43.08 (13.72)	43.43 (15.69)	47.21 (14.96)
Age range	18-69	20-84	20-74
GCS Mean (SD)	7.11 (4.07)	8.0 (4.06)	7.92 (2.87)
Weeks post-injury Mean (SD)	76.03 (136.15)	118.17 (182.26)	62.68 (124.83)
Weeks post-injury range	4-525	2-550	5-525
SPANS Total Score	185.58 (47.22)	200.28 (48.45)	204.26 (47.16)

Overall, the university educated sample had a slightly higher mean age than the other samples, whereas the college/vocational level sample had a much higher mean weeks post-injury. Mean SPANS score increased with level of education. ROC curves are displayed in Figures 6, 7, 8, 9, 10 and 11 and AUC and significance levels in Table 9, 10 and 11.

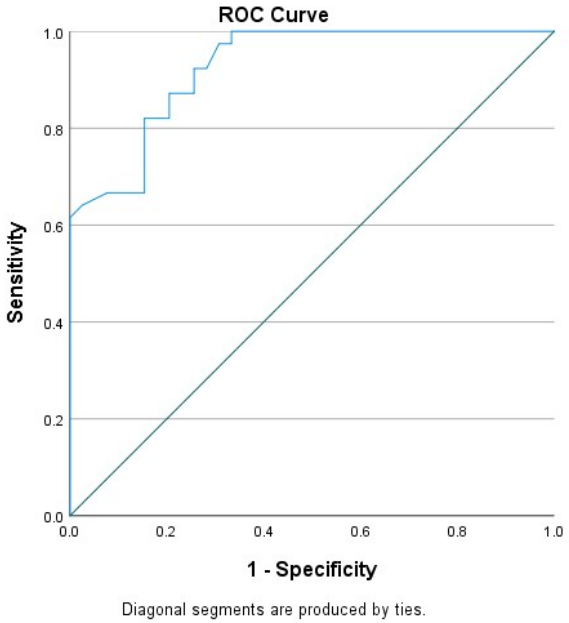

Figure 6: ROC curve for SPANS total score in high school educated sample

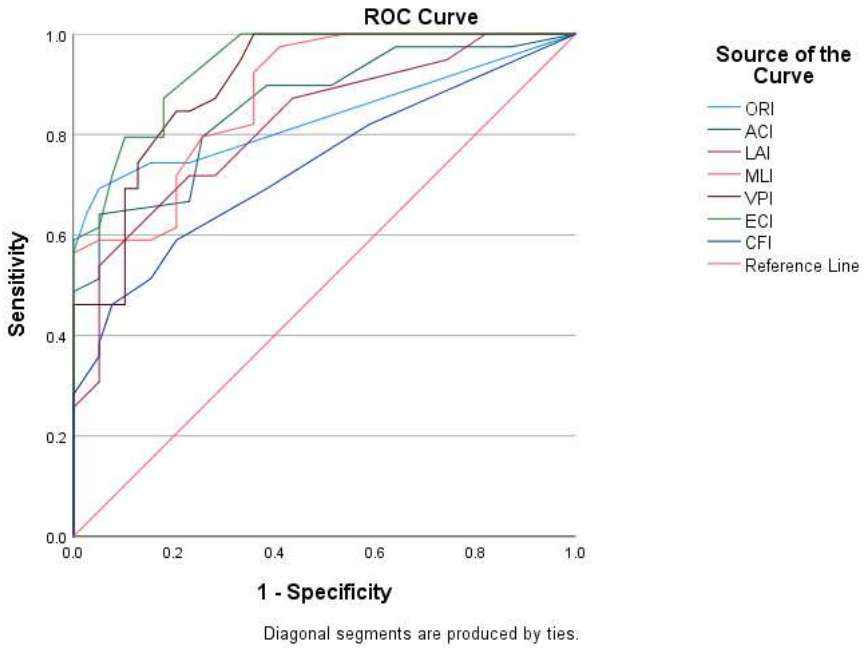

Figure 7: ROC curves for SPANS index scores in high school educated sample

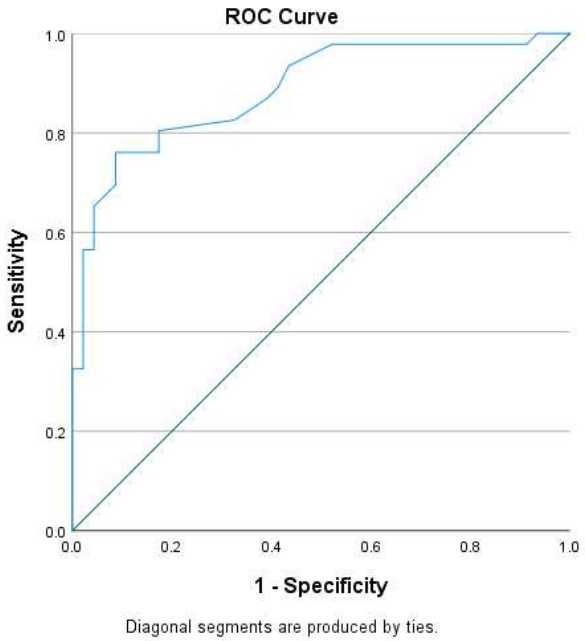

Figure 8: ROC curve for SPANS total score in college/vocational educated sample

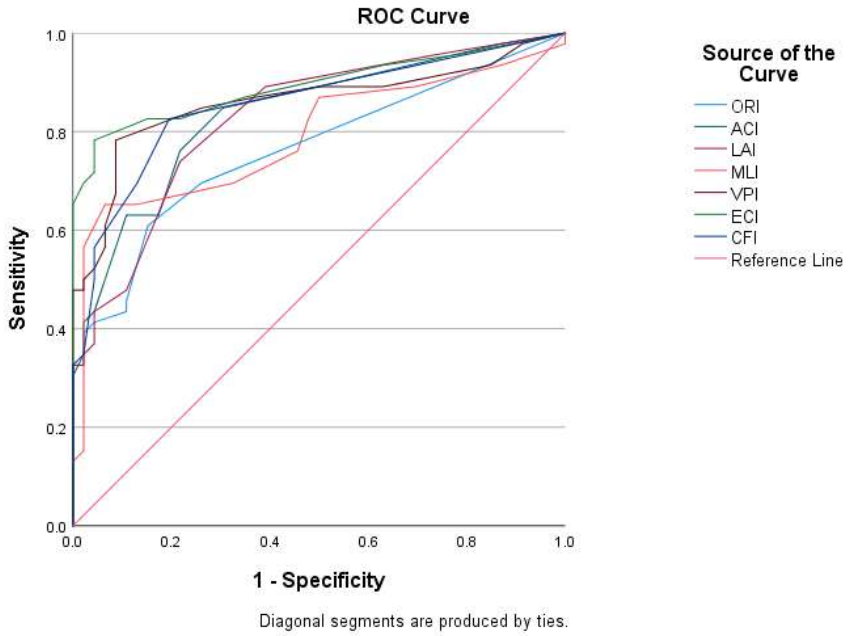

Figure 9: ROC curves for SPANS index scores in college/vocational educated sample

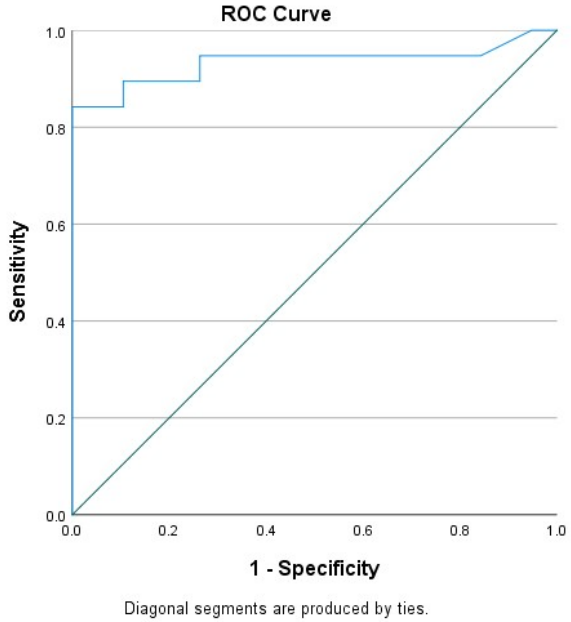

Figure 10: ROC curve for SPANS total score in university educated sample

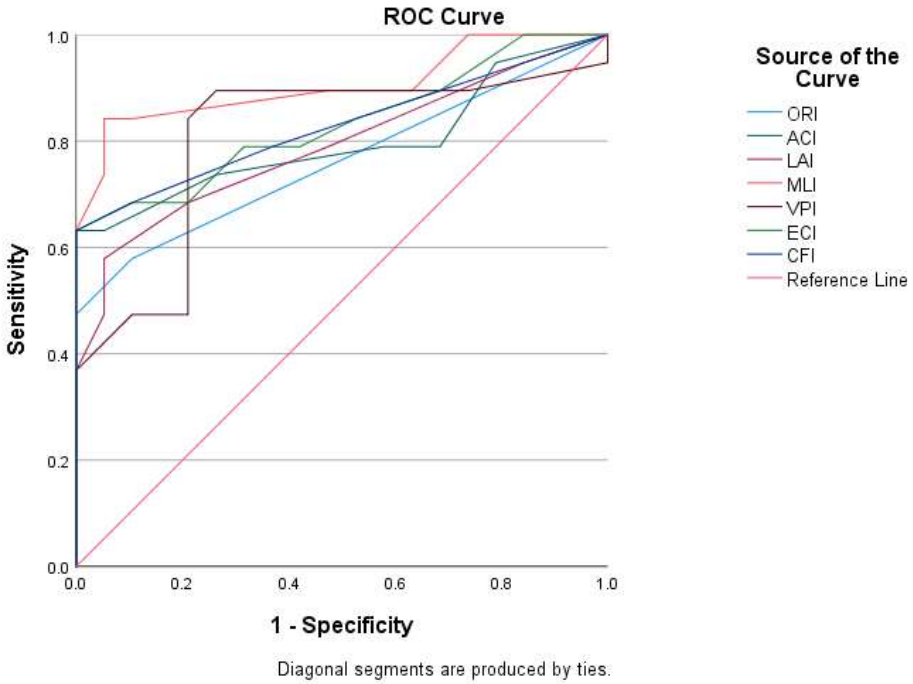

Figure 11: ROC curves for SPANS index scores in university educated sample

Table 9: AUC statistic for SPANS total score and indices for high school educated sample

				95% Confidence Interval	
Test Variable	AUC	Std. Error	Significance	Lower Bound	Upper Bound
SPANS total	.927	.027	.000	.875	.980
ORI	.834	.049	.000	.738	.930
ACI	.857	.042	.000	.774	.939
LAI	.821	.047	.000	.729	.913
MLI	.879	.037	.000	.808	.951
VPI	.905	.033	.000	.840	.969
ECI	.939	.024	.000	.892	.986
CFI	.737	.057	.000	.625	.848

Table 10: AUC statistics for SPANS total score and indices for college/vocational sample

				95% Confidence Interval	
Test Variable	AUC	Std. Error	Significance	Lower Bound	Upper Bound
SPANS total	.886	.035	.000	.818	.954
ORI	.764	.050	.000	.665	.863
ACI	.837	.042	.000	.755	.919
LAI	.829	.043	.000	.746	.913
MLI	.800	.049	.000	.694	.887
VPI	.860	.042	.000	.778	.943
ECI	.893	.036	.000	.822	.965
CFI	.855	.041	.000	.776	.935

Table 11: AUC statistics for SPANS total score and indices scores for University educated sample

				95% Confidence Interval	
Test Variable	AUC	Std. Error	Significance	Lower Bound	Upper Bound
SPANS total	.934	.048	.000	.839	1.000
ORI	.762	.081	.006	.604	.920
ACI	.798	.078	.002	.645	.950
LAI	.789	.076	.002	.641	.938
MLI	.898	.053	.000	.800	1.000
VPI	.806	.077	.001	.656	.957
ECI	.832	.068	.000	.698	.966
CFI	.799	.071	.001	.689	.96

As shown in Figures 5, 6 and 7, ROC curves for all three education levels were above the line of no-interest and towards the left hand 'perfect' discrimination corner of the axis. AUC's for the high school and University educated samples were outstanding, while the college/vocational education level was excellent and approached outstanding. All AUC's were highly significant ($p < .001$) indicating that the SPANS had good discriminative ability across all education levels. ROC curves for all individual indices were also towards the left-hand corner, supported by AUCs ranging from acceptable to outstanding. Individual index AUCs were highest for the high school educated sample and lowest for the university educated sample, which is likely due to the much lower *N* for this sample, although all were still at least 'acceptable' discriminators.

As above classification statistics were calculated and two cut-off scores were chosen, one which maximised sensitivity and one that maximised specificity. A third cut-off score, where sensitivity/specificity were given equal weight was also chosen, where possible, to

make comparisons between cut-offs easier. Classification statistics for the SPANS total score for the three different education samples are displayed in Tables 12, 13 and 14.

Table 12: Classification statistics for the SPANS total score and indices in high school educated ABI sample

Index	Prioritised	Cut-off	Sensitivity	Specificity	Youden's	+LR	-LR	PPV	NPV
SPANS Total*	Sensitivity	233	92.3%	74.5%	.668	3.615	0.103	76%	88%
	Specificity	216	66.7%	91.5%	.582	7.833	0.364	96%	73%
	Equal	230	87.2%	80.9%	.680	4.552	0.158	80%	82%
ORI*	Sensitivity	21	74.4%	85.1%	.595	4.992	0.301	90%	75%
	Specificity	20	64.1%	95.7%	.628	10.846	0.328	100%	70%
	Equal	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ACI*	Sensitivity	41	89.7%	63.8%	.536	2.481	0.160	76%	78%
	Specificity	38	64.1%	93.6%	.577	10.042	0.383	88%	68%
	Equal	40	79.5%	74.5%	.540	3.113	0.275	76%	70%
LAI*	Sensitivity	50	87.2%	55.3%	.425	1.951	0.231	68%	70%
	Specificity	47	66.7%	80.9%	.475	3.481	0.412	91%	67%
	Equal	48	71.8%	76.6%	.484	3.067	0.368	76%	70%
MLI*	Sensitivity	60	92.3%	66.0%	.583	2.711	0.116	71%	79%
	Specificity	57	71.8%	80.9%	.526	3.749	0.348	75%	67%
	Equal	58	79.5%	76.6%	.561	3.396	0.267	80%	74%
VPI*	Sensitivity	63	94.9%	66.0%	.608	2.786	0.077	76%	93%
	Specificity	57	69.2%	91.5%	.607	8.134	0.336	87%	74%
	Equal	60	84.6%	78.7%	.633	3.976	0.195	85%	77%
ECI*	Sensitivity	40	87.2%	76.6%	.638	3.724	0.167	79%	79%
	Specificity	37	71.8%	91.5%	.633	8.435	0.308	92%	71%
	Equal	38	79.5%	87.2%	.667	6.226	0.235	90%	77%
CFI*	Sensitivity	27	82.1%	40.4%	.225	1.377	0.443	61%	65%
	Specificity	25	59.0%	78.7%	.377	2.771	0.521	74%	63%
	Equal	26	69.2%	61.7%	.309	1.807	0.498	70%	64%

*Maximum total scores: SPANS Total = 334, ORI = 22, ACI = 46, LAI = 53, MLI = 67, VPI = 70, ECI = 48, CFI = 28

Table 13: Classification statistics for the SPANS total score and individual indices in college/vocational educated ABI sample

Index	Prioritised	Cut-off	Sensitivity	Specificity	Youden's	+LR	-LR	PPV	NPV
SPANS Total*	Sensitivity	242	80.4%	76.9%	.574	3.485	0.254	80%	77%
	Specificity/Equal	237	76.1%	90.4%	.665	7.913	0.264	89%	76%
ORI*	Sensitivity	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Specificity	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Equal	21	60.9%	82.7%	.436	3.516	0.473	78%	62%
ACI*	Sensitivity	43	84.8%	63.7%	,521	2.593	0.226	76%	76%
	Specificity	40	63.0%	88.5%	.515	5.463	0.417	85%	65%
	Equal	42	76.1%	76.9%	.530	3.297	0.310	74%	68%
LAI*	Sensitivity	51	89.1%	57.7%	.468	2.106	0.188	81%	76%
	Specificity	49	47.8%	88.5%	.363	4.144	0.589	98%	68%
	Equal	50	73.9%	78.8%	.528	3.494	0.330	88%	64%
MLI*	Sensitivity	62	82.6%	53.8%	.365	1.789	0.322	63%	69%
	Specificity	58	65.2%	92.3%	.575	8.478	0.376	93%	69%
	Equal	60	69.6%	69.2%	.388	2.260	0.439	83%	71%
VPI*	Sensitivity	65	87.0%	59.6%	.466	2.153	0.218	76%	83%
	Specificity	63	78.3%	92.3%	.706	10.173	0.235	91%	74%
	Equal	64	84.8%	75.0%	.598	3.391	0.202	92%	81%
ECI*	Sensitivity	45	87.0%	65.4%	.523	2.512	0.199	79%	81%
	Specificity	42	78.3%	92.3%	.706	10.173	0.235	92%	80%
	Equal	43	82.6%	82.7%	.653	4.772	0.210	90%	81%
CFI*	Sensitivity	27	82.6%	76.9%	.595	3.579	0.226	86%	75%
	Specificity	25	56.5%	94.2%	.508	9.797	0.461	92%	66%
	Equal	26	69.6%	86.5%	.561	5.167	0.351	93%	69%

*Maximum total scores: SPANS Total = 334, ORI = 22, ACI = 46, LAI = 53, MLI = 67, VPI = 70, ECI = 48, CFI = 28

Table 14: Classification statistics for the SPANS total score and individual indices in university educated ABI sample

Index	Prioritised	Cut-off	Sensitivity	Specificity	Youden's	+LR	-LR	PPV	NPV
SPANS Total*	Sensitivity	241	94.7%	80.0%	.747	4.736	0.065	89%	89%
	Specificity	231	84.2%	100%	.842	N/A	0.157	100%	86%
	Equal	237	89.5%	92.0%	.815	11.184	0.114	89%	85
ORI*	Sensitivity	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Specificity	21	57.9%	88.0%	.459	4.824	0.478	100%	66
	Equal	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ACI*	Sensitivity	43	78.9%	52.0%	,309	1.644	0.404	78%	75%
	Specificity	38	63.2%	100%	.632	N/A	0.368	100%	68%
	Equal	42	73.7%	80.0%	.537	3.684	0.328	92%	72%
LAI*	Sensitivity	51	78.9%	60.0%	.389	1.973	0.350	72%	70%
	Specificity/equal	50	68.4%	80.0%	.484	3.421	0.394	92%	69%
MLI*	Sensitivity	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Specificity	54	63.2%	100%	.632	N/A	0.368	100%	73%
	Equal	60	84.2%	96.0%	.802	21.052	0.164	100%	83%
VPI*	Sensitivity	64	89.5%	80.0%	.682	5.263	0.187	89%	85%
	Specificity	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Equal	N/A	N/A	N/A	N/A	N/A	N/A	89%	89%
ECI*	Sensitivity	45	84.2%	60.0%	.442	2.105	0.263	75%	78%
	Specificity	40	68.4%	92.0%	.604	8.552	0.343	100%	73%
	Equal	43	78.9%	76.0%	.549	3.289	0.277	87%	74%
CFI*	Sensitivity	27	78.9%	68.0%	.469	2.467	0.309	81%	73%
	Specificity	25	63.2%	100%	.632	N/A	0.368	100%	70%
	Equal	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

*Maximum total scores: SPANS Total = 334, ORI = 22, ACI = 46, LAI = 53, MLI = 67, VPI = 70, ECI = 48, CFI = 28

Cut-offs for the SPANS total score, which exceeded Lincoln et al.'s (2003) criteria for a diagnostic test, were identified for all three levels of education. The high school educated sample and university educated sample had particularly high sensitivity/specificity with both producing different cut-offs with sensitivities of over 90% and specificity levels approaching or exceeding 80%. The university educated sample demonstrated an outstanding 'perfect' specificity of 1 whilst maintaining a sensitivity level of over 80% indicating extremely good diagnostic validity in this population. Sensitivity and specificity levels for participants with college or vocational education were lower but still produced cut-offs with sensitivity over 80% and specificity over 75%. Lower levels of sensitivity and specificity were found for the

individual index cut-offs, with the ECI being the only index producing adequate levels of sensitivity and specificity for a diagnostic test (Lincoln et al., 2003) across all three samples.

Optimal cut-off scores for the SPANS total score for both college/vocational and university educated samples were very similar with cut-offs prioritising sensitivity and those giving equal prioritisation to both sensitivity and specificity being within 1 point of each other for both samples. The cut-off which gave equal prioritisation to both for the college/vocational education sample was the same cut-off which prioritised specificity. Cut-off scores for the sample of participants with high school level education were significantly lower than the other education levels for all three prioritisation criteria. Cut-offs prioritising sensitivity and giving equal prioritisation were approximately 8 points lower for this sample whilst the cut-off prioritising specificity was over 15 points lower. This trend continued at an individual index level with cut-offs for all indices apart from the ORI and CFI being higher in the college/vocational and university educated samples. The ECI appeared to show the most difference in optimal cut-offs between these samples.

4 Discussion

The current study assessed the ability of the SPANS total and individual index scores to discriminate between brain-injured and healthy norm participants in acute and long-term ABI settings and for three different education levels. Results provide evidence that the SPANS is a valid neuropsychological screen in both acute (less than 6 months post injury) and longer-term (over 1-year post injury) ABI populations. Education level significantly impacted on classification statistics.

4.1 Discriminant ability of SPANS scores in acute and long-term ABI populations

4.1.1 SPANS Total score

The AUC statistic for the SPANS total score in the acute population neared perfect discrimination between brain-injured and healthy norm samples. As hypothesised the AUC for the SPANS total score was slightly lower in the long-term sample, which would be expected given that cognitive impairments at this time-point are likely to be much milder due to the steep recovery in the first few months post-injury (Christensen et al., 2008), this is supported by the higher mean SPANS total score for the longer-term sample. However further analysis indicated that differences between the AUCs in the two samples were not statistically significant, meaning it cannot be concluded that the SPANS is a better discriminator in the acute sample from this analysis. The AUC for the acute sample still indicated 'excellent' discrimination between the ABI sample and healthy controls indicating the SPANS is a valid global cognitive screen in both populations.

4.1.2 SPANS Index Scores

All individual SPANS indices showed excellent-to-outstanding AUC's in the acute population, as hypothesised, indicating that each individual index was able to successfully discrimin/ate between people who have suffered an ABI and those who have not. The VPI and ECI were the best discriminators with 'outstanding' AUCs of over 90% and the ECI alone exceeding the AUC for the overall SPANS total score. Statistical analysis showed that the difference between the VPI and ECI AUCs and those of the other indices was statistically significant. This could be evidence of these cognitive domains being the most impaired in the acute phase post-ABI. The fact that each index achieved at least excellent AUC's supports the theoretical construct of the SPANS and indicates that impairments in all domains chosen to be included in the SPANS were present and relevant to acute ABI patients. These results support the discriminative ability of the SPANS at an index level, providing initial evidence of its utility as a domain-specific screen, in this sample. Further ROC analysis, with samples where impairment in each domain has been established would be needed to confirm this.

The ability of the individual SPANS indices to discriminate between long-term ABI patients and healthy norms was lower and varied between indices, as hypothesised. This could indicate lower levels of impairment and different patterns of recovery across cognitive domains or be due to a lack of sensitivity in SPANS indices. Without further ROC analysis in this population, with domain-specific impairment established before inclusion, this cannot be known. The best discriminators in the long-term sample were the ECI and MLI, which both had 'excellent' AUC statistics which were significantly different to those of the other indices. This is in line with hypotheses based on previous research that memory impairments endure long-term post-ABI (Vanderploeg et al., 2005), indicating that these impairments should be a focus of rehabilitation plans and resource allocation. As hypothesised the ACI was still an acceptable discriminator in the long-term population lending further support to the fact that impairments in attention are maintained long-term (Vanderploeg et al., 2005). The fact that this index had lower discriminative ability that the MLI could indicate that enduring impairments in attention are not as severe as those in the domain of memory. Alternatively, it could be because of a lack of sensitivity in the SPANS attention index, further analysis would be needed to clarify.

The LAI ,and VPI were also 'acceptable' discriminators, despite previous research showing that language and dexterity tend to recover (Christensen et al., 2008; Shiel et al., 2000). This indicates that attention, language, visuospatial and processing speed impairments were still present in a large proportion of the long-term sample. It is possible that the SPANS indices were more sensitive to these impairments than the subtests used in previous research, or, results could have been influenced by the sampling method for this study, as all participants were referred to neuropsychology for assessment and so were experiencing potential issues with their cognitive functioning. This could mean that levels of impairment

were artificially high compared to the general ABI population., though again, this cannot be confirmed without further ROC analysis.

The ORI and CFI both had AUCs classed as 'poor' however, further analysis indicated that differences in these AUCs and those of the other indices were not statistically significant, except for in the case of the MLI and ECI meaning it cannot be concluded that these indices were poorer discriminators than the ACI, LAI and VPI. The fact that these indices still discriminated at an above chance level indicates that although these impairments may show steep recovery, they are still present in some long-term brain-injured patients. This also provides evidence that the ORI and CFI may have high sensitivity and were therefore able to pick up the more subtle impairments that may be present at this time-point.

4.1.3 Classification Statistics

Classification statistics showed that the SPANS total score had high levels of sensitivity and specificity in both longer-term and acute ABI populations, indicating that it is sensitive enough to detect the type and level of cognitive impairment that may persist beyond a year, while specific enough to distinguish between the ABI sample, and healthy norm. As expected, the SPANS total cut-off score was higher (~3%) in the long-term sample, suggesting this was necessary in order to achieve adequate sensitivity due to the less severe cognitive impairments present. Cut-offs were produced for the SPANS total score for both samples that provided good levels of sensitivity/specificity levels meeting Lincoln et al.'s (2003) criteria, suggesting that the SPANS is an acceptable diagnostic test of injury to the brain. The fact that these high levels were produced for different cut-off values supports the need to adjust cut-offs based on time-since-injury.

In the acute population, sensitivity/specificity was high across all indices with each producing cut-offs meeting Lincoln et al.'s (2003) criteria. As expected, sensitivity/specificity levels were lower and more varied for individual indices in the longer-

term sample, with only the MLI having a cut-off score able to meet Lincoln et al.'s criteria for a diagnostic test. The ORI and CFI cut-offs had particularly poor sensitivity in this sample, providing support for the hypothesis that these would be poor discriminators based on previous research indicating these domains show good recovery (e.g. Christenson et al., 2008), though again this cannot be confirmed with the current study design.

For the SPANS total score, and most indices, it was possible to select two cut-off scores prioritising sensitivity and specificity respectively. These cut-offs produced very different likelihood ratios and predictive values indicating differing clinical utility, with more sensitive cut-offs providing fewer false negatives and more specific cut-offs fewer false positives. These findings provide support for the provision of separate scores which can be selected according to the clinical need.

4.2 Impact of education level on optimal cut-off scores

Results supported the hypothesis and previous research findings that education level impacts scores on cognitive assessments (Malek-Ahmadi et al., 2015). The AUCs for the SPANS total score were outstanding for the high school and university educated samples and excellent for the college/vocational education sample, showing that the SPANS could successfully discriminate between ABI patients and healthy norms regardless of education level. The college or vocational education sample used in the analysis had a much higher average weeks post-injury (see Table 8) meaning that impairments were likely to be milder in this sample, perhaps partially accounting for the lower AUC.

As hypothesised, optimal cut-off scores for the SPANS total score varied significantly between samples. Cut-off scores for participants with some college or vocational education and participants with university level education were roughly equivalent. Optimal cut-offs selected for the sample of participants with high school education however were much lower, between 8-15 points (~5%), depending on the prioritisation given to sensitivity or specificity.

When this was investigated at an index level, the pattern of higher cut-offs persisted across all indices, apart from the ORI and CFI. This suggests that orientation and conceptual flexibility are skills that are not generally impacted by level of education.

Very high levels of sensitivity and specificity were found for all three samples, based on different cut-off values, highlighting the importance of these cut-offs for the validity of the SPANS as a diagnostic test. Classification statistics demonstrated that using the higher scores indicated in the samples with higher education would have poor levels of sensitivity for patients with lower education levels, resulting in many missed cases of impairment. On the other hand, if the lower cut-offs indicated for the lower levels of education were used globally then this would result in more patients with higher levels of education being falsely classified as impaired.

4.3 Limitations of the study

The current study had several limitations, mainly due to the retrospective nature of the sample. Firstly, participants were not consecutively recruited, and therefore not representative of the general ABI population. The fact that participants were only included after being referred for neuropsychological testing could have produced higher AUCs, as to be referred they would likely be having problems with their cognition. However, as participants were gathered as part of everyday clinical practice with only those for whom SPANS results would be relevant being recruited, it could be argued that this method of selection produces a more ecologically valid sample for the validation of the SPANS. Additionally, the data contained assessment results from both SPANS versions (A & B) combined. Due to limited sample sizes, and evidence of the equivalency of versions (Burgess, 2014) data from both versions were included in the same analyses to maintain power. However, the different SPANS forms may have different levels of discriminant validity and different potential cut-offs. The inclusion of re-test data for some analysis which could have biased results. However, ROCs

run without these data included did not show significantly different results and the mean SPANS scores were significantly different between these time-points suggesting distinct cognitive presentations.

The main limitation of the study was the lack of confirmation of impairment, with a gold-standard test, before data were used for ROC analysis. This creates problems for the interpretation of results and the ability to draw conclusions on the discriminant ability of the SPANS, particularly for the post one-year since injury analysis. As discussed above, it is likely that many areas of impairment have improved by this timepoint (e.g. Sigurdardottir et al., 2020). Even in the acute sample it is not realistic to conclude that every person who has suffered an ABI is impaired in each of the cognitive domains assessed by the SPANS indices. assessed. Without the confirmation of impairment, it is impossible to ascertain whether levels of sensitivity/specificity are accurate or are underestimations of its true diagnostic abilities.

4.4 Clinical implications

4.4.1 Utility of the SPANS

The current study provides evidence that the SPANS is a sensitive and specific global cognitive screen for ABI in both acute and long-term samples. Results also indicate that each individual SPANS index is a sensitive and specific diagnostic tool for acute ABI. This, together with previous evidence of the internal reliability and convergent validity of its indices, suggests it is a valid and 'sufficient' domain-specific cognitive screen for use in acute ABI patients. Some evidence is provided for SPANS indices being sensitive diagnostic tools in the longer-term populations, with indices found to be 'acceptable' discriminators even for domains which were expected to recover, although levels of sensitivity and specificity for individual cut-offs was low. As the actual level of impairment for each participant in each cognitive domain was unknown, no firm conclusions can be drawn for the validity of the SPANS as a domain-specific cognitive screen in long-term ABI. It is likely

that participants from the clinical sample had recovered functioning in many of the cognitive domains at this point meaning results may have underestimated the diagnostic ability of the SPANS.

4.4.2 Recovery of cognitive domains post-ABI

Results from this study indicate that memory and efficiency/cognitive flexibility were the cognitive domain with the most severe and enduring impairments at one-year post-ABI, suggesting that they should be a focus of rehabilitation plans and provision of care. This study also found that impairments in visuomotor, attention and language were still present in many post-one-year ABI patients. Additionally, although orientation and conceptual flexibility had recovered, as predicted, they were still detectable in the long-term sample. The sampling method for this study may have resulted in participants having higher than average cognitive impairment, due to them having been referred for cognitive testing. However, this is still an indication that these impairments do endure long-term for a proportion of people post-ABI. Rehabilitation and care plans should therefore make provisions for a range of impairments and not assume full recovery of function.

4.4.3 Impact of participant factors and level of sensitivity/specificity on cut-off scores

Results from this study indicate that level of education and time-since injury significantly impact on optimal cut-off scores for neuropsychological tests. Extremely high levels of sensitivity and specificity were able to be found for samples of people with both lower and higher education levels, however, these were found for very different cut-off scores and would therefore not be the case if a blanket cut-off score was selected. Additionally, for both acute and long-term samples, cut-offs were able to be provided which maximised either sensitivity or specificity, producing very different classification statistics.

Overall, these results suggest that the selection of one unitary cut-off for all patients, regardless of patient variables or clinical need, is reductive and likely significantly reduces

the validity of cognitive assessments. It is likely that many factors, other than those investigated here, also have an impact on cognitive assessment scores. An over-reliance on unitary cut-offs and sensitivity/specificity estimates, often produced through analysis of one population sample, is therefore misleading and is likely to result in cases of impairment being missed, and others being falsely diagnosed. Instead, a more nuanced approach is indicated with the production of cut-off scores able to be adjusted to relevant patient characteristics and to the level of sensitivity/specificity best fitted to the clinical need. Even in this case, arbitrary cut-offs should be treated with caution at an individual level.

5 References

Bedard, M., Steffener, J., & Taler, V. (2020). Long-term cognitive impairment following single mild traumatic brain injury with loss of consciousness: Findings from the Canadian Longitudinal Study on Aging. *Journal of Clinical and Experimental Neuropsychology, 42*(4), 344-351.

Burgess, G. (2014). *Short Parallel Assessments of Neuropsychological Status: SPANS mannual*. Oxford: Hogrefe.

Carroll, E. L., Outtrim, J. G., Forsyth, F. et al. (2020). Mild traumatic brain injury recovery: a growth curve modelling analysis over 2 years. *Journal of Neurology*, XX, 1-12.

Cartoni, A., & Lincoln, N. (2005). The sensitivity and specificity of the Middlesex Elderly Assessment of Mental State (MEAMS) for detecting cognitive impairment after stroke. *Neuropsychological rehabilitation, 15*(1), 55-67.

Chiti, G., & Pantoni, L. (2014). Use of Montreal Cognitive Assessment in patients with stroke. *Stroke, 45*(10), 3135-3140.

Christensen, B. K., Colella, B., Inness, E. et al. (2008). Recovery of cognitive function after traumatic brain injury: a multilevel modeling analysis of Canadian outcomes. *Archives of Physical Medicine and Rehabilitation, 89*(12), S3-S15.

Fleminger, S., & Ponsford, J. (2005). Long term outcome after traumatic brain injury: British Medical Journal Publishing Group.

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.

Fortune, D. G., Walsh, R. S., & Richards, H. L. (2016). Cognitive reserve and preinjury educational attainment: effects on outcome of community-based rehabilitation for longer-term individuals with acquired brain injury. *International Journal of Rehabilitation Research, 39*(3), 234-239.

Greenwald, B. D., Burnett, D. M., & Miller, M. A. (2003). Congenital and acquired brain injury. 1. Brain injury: epidemiology and pathophysiology. *Archives of Physical Medicine and Rehabilitation, 84*(3 Suppl 1), S3-S7.

Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression, John Wiley & Sons. *New York*.

Lincoln, N., Nicholl, C., Flannaghan, T. et al. (2003). The validity of questionnaire measures for assessing depression after stroke. *Clinical rehabilitation, 17*(8), 840-846.

Malek-Ahmadi, M., Powell, J. J., Belden, C. M. et al. (2015). Age- and education-adjusted normative data for the Montreal Cognitive Assessment (MoCA) in older adults age 70–99. *Aging, Neuropsychology, and Cognition, 22*(6), 755-761.

Mathuranath, P., Cherian, P., Mathew, R. et al. (2006). P2–086: The Addenbrooke's Cognitive Examination: Norms and validation for a multicultural Indian population. *Alzheimer's & Dementia, 2*, S258-S258.

McCauley, S. R., Wilde, E. A., Barnes, A. et al. (2014). Patterns of early emotional and neuropsychological sequelae after mild traumatic brain injury. *Journal of Neurotrauma, 31*(10), 914-925.

Nabors, N. A., Millis, S. R., & Rosenthal, M. (1997). Use of the Neurobehavioral Cognitive Status Examination (Cognistat) in traumatic brain injury. *The Journal of Head Trauma Rehabilitation*.

Narasimhalu, K., Ang, S., De Silva, D. et al. (2009). Severity of CIND and MCI predict incidence of dementia in an ischemic stroke cohort. *Neurology, 73*(22), 1866-1872.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V. et al. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699.

Schretlen, D. J., & Shapiro, A. M. (2003). A quantitative review of the effects of traumatic brain injury on cognitive functioning. *International review of psychiatry, 15*(4), 341-349.

Shiel, A., Horn, S., Wilson, B., Watson, M. et al. (2000). The Wessex Head Injury Matrix (WHIM) main scale: a preliminary report on a scale to assess and monitor patient recovery after severe head injury. *Clinical rehabilitation, 14*(4), 408-416.

Sigurdardottir, S., Andelic, N., Roe, C. et al. (2009). Cognitive recovery and predictors of functional outcome 1 year after traumatic brain injury. *Journal of the International Neuropsychological Society: JINS, 15*(5), 740.

Sigurdardottir, S., Andelic, N., Røe, C. et al. (2020). Trajectory of 10-year neurocognitive functioning after moderate–severe traumatic brain injury: Early associations and clinical application. *Journal of the International Neuropsychological Society*.

Smania, N., Avesani, R., Roncari, L. et al. (2013). Factors predicting functional and cognitive recovery following severe traumatic, anoxic, and cerebrovascular brain damage. *The Journal of Head Trauma Rehabilitation, 28*(2), 131-140.

Teasdale, G., & Jennett, B. (1974). The Glasgow coma scale. *Lancet, 2*, 81-84.

Vanderploeg, R. D., Curtiss, G., & Belanger, H. G. (2005). Long-term neuropsychological outcomes following mild traumatic brain injury. *Journal of the International Neuropsychological Society: JINS, 11*(3), 228.

Weicker, J., Villringer, A., & Thöne-Otto, A. (2016). Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. *Neuropsychology, 30*(2), 190-212.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32-35.

Zinn, S., Dudley, T. K., Bosworth, H. B. et al. (2004). The effect of poststroke cognitive impairment on rehabilitation process and functional outcome. *Archives of Physical Medicine and Rehabilitation, 85*(7), 1084-1090.

Appendices

Appendix A: Chosen journal formatting guideline for Literature Review

The Journal of Neuropsychological Rehabilitation was selected as the target journal for the literature review. Author guidelines are as follows:

Preparing Your Paper

All authors submitting to medicine, biomedicine, health sciences, allied and public health journals should conform to the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, prepared by the International Committee of Medical Journal Editors (ICMJE).

Clinical trials: must conform to the Consort guidelines http://www.consort-statement.org. Submitted papers should include a checklist confirming that all of the Consort requirements have been met, together with the corresponding page number of the manuscript where the information is located. In addition, trials must be pre-registered on a site such as clinicaltrials.gov or equivalent, and the manuscript should include the reference number to the relevant pre-registration.

Systematic reviews: submitted papers should follow PRISMA http://www.prisma-statement.org/ guidelines and submission should also be accompanied by a completed PRISMA checklist, together with the corresponding page number of the manuscript where the information is located.

Single-case studies: submitted papers should follow SCRIBE guidelines (http://psycnet.apa.org/fulltext/2016-17384-001.html) and include a completed SCRIBE checklist together with the corresponding page number of the manuscript where the information is located.

Observational studies: submitted papers should follow the STROBE guidelines (https://www.strobe-statement.org/index.php?id=strobe-home) and also include a completed checklist of compliance, together with the corresponding page number of the manuscript where the information is located.

Qualitative studies: should follow the COREQ guidelines (http://www.equator-network.org/reporting-guidelines/coreq/) and be accompanied by a completed COREQ checklist of compliance, together with the corresponding page number of the manuscript where the information is located.

The EQUATOR Network (Enhancing the Quality and Transparency of Health Research) website provides further information on available guidelines.

Structure

Your paper should be compiled in the following order: title page; abstract; keywords; main text introduction, materials and methods, results, discussion; acknowledgments; declaration of interest statement; references; appendices (as appropriate); table(s) with caption(s) (on individual pages); figures; figure captions (as a list).

Word Limits

Please include a word count for your paper. There are no word limits for papers in this journal.

Format-Free Submission

Authors may submit their paper in any scholarly format or layout. Manuscripts may be supplied as single or multiple files. These can be Word, rich text format (rtf), open document format (odt), or PDF files. Figures and tables can be placed within the text or submitted as separate documents. Figures should be of sufficient resolution to enable refereeing.

- There are no strict formatting requirements, but all manuscripts must contain the essential elements needed to evaluate a manuscript: abstract, author affiliation, figures, tables, funder information, and references. Further details may be requested upon acceptance.
- References can be in any style or format, so long as a consistent scholarly citation format is applied. Author name(s), journal or book title, article or chapter title, year of publication, volume and issue (where appropriate) and page numbers are essential. All bibliographic entries must contain a corresponding in-text citation. The addition of DOI (Digital Object Identifier) numbers is recommended but not essential.
- The journal reference style will be applied to the paper post-acceptance by Taylor & Francis.
- Spelling can be US or UK English so long as usage is consistent.

Note that, regardless of the file format of the original submission, an editable version of the article must be supplied at the revision stage.

Checklist: What to Include

1. **Author details.** Please ensure everyone meeting the International Committee of Medical Journal Editors (ICMJE) requirements for authorship is included as an author of your paper. All authors of a manuscript should include their full name and affiliation on the cover page of the manuscript. Where available, please also include ORCiDs and social media handles (Facebook, Twitter or LinkedIn). One author will need to be identified as the corresponding author, with their email address normally displayed in the article PDF (depending on the journal) and the online article. Authors' affiliations are the affiliations where the research was conducted. If any of the named co-authors moves affiliation during the peer-review process, the new affiliation can be given as a footnote. Please note that no changes to affiliation can be made after your paper is accepted. Read more on authorship.
2. Should contain an unstructured abstract of 200 words.
3. You can opt to include a **video abstract** with your article. Find out how these can help your work reach a wider audience, and what to think about when filming.
4. Between 5 and 5 **keywords**. Read making your article more discoverable, including information on choosing a title and search engine optimization.
5. **Funding details.** Please supply all details required by your funding and grant-awarding bodies as follows:
   For single agency grants

This work was supported by the [Funding Agency] under Grant [number xxxx].
For multiple agency grants
This work was supported by the [Funding Agency #1] under Grant [number xxxx]; [Funding Agency #2] under Grant [number xxxx]; and [Funding Agency #3] under Grant [number xxxx].

6. **Disclosure statement.** This is to acknowledge any financial interest or benefit that has arisen from the direct applications of your research. Further guidance on what is a conflict of interest and how to disclose it.
7. **Data availability statement.** If there is a data set associated with the paper, please provide information about where the data supporting the results or analyses presented in the paper can be found. Where applicable, this should include the hyperlink, DOI or other persistent identifier associated with the data set(s). Templates are also available to support authors.
8. **Data deposition.** If you choose to share or make the data underlying the study open, please deposit your data in a recognized data repository prior to or at the time of submission. You will be asked to provide the DOI, pre-reserved DOI, or other persistent identifier for the data set.
9. **Geolocation information.** Submitting a geolocation information section, as a separate paragraph before your acknowledgements, means we can index your paper's study area accurately in JournalMap's geographic literature database and make your article more discoverable to others. More information.
10. **Supplemental online material.** Supplemental material can be a video, dataset, fileset, sound file or anything which supports (and is pertinent to) your paper. We publish supplemental material online via Figshare. Find out more about supplemental material and how to submit it with your article.
11. **Figures.** Figures should be high quality (1200 dpi for line art, 600 dpi for grayscale and 300 dpi for colour, at the correct size). Figures should be supplied in one of our preferred file formats: EPS, PS, JPEG, TIFF, or Microsoft Word (DOC or DOCX) files are acceptable for figures that have been drawn in Word. For information relating to other file types, please consult our Submission of electronic artwork document.
12. **Tables.** Tables should present new information rather than duplicating what is in the text. Readers should be able to interpret the table without reference to the text. Please supply editable files.
13. **Equations.** If you are submitting your manuscript as a Word document, please ensure that equations are editable. More information about mathematical symbols and equations.
14. **Units.** Please use SI units (non-italicized).

Appendix B: Data extraction form used in Literature review

Authors	Location and setting (country and context e.g. stroke ward)	Sample (Including sampling method, number of participants, demographic, medical and cognitive descriptors)	Cognitive 'domains' used as Predictors	Cognitive assessment tools used, and author defined labels	Assessment time points (for cognitive and functional variables)	Author-defined functional outcomes assessed, and assessment tools used	Statistical Analysis (type of analysis, all reported statistics, variables included and controlled for)	Potential Confounds (including whether these were controlled for)	Key Findings

Appendix **C: Summary of all included studies**

Study	**Cognitive domains and assessments**	**Outcome categories and measures**	**Assessment time-points**	**Type of statistical analysis**
Bertolin et al., 2018	Executive Function (TMT A & B)	Activities of Daily Living (SIS subscales) Participation (SIS subscales)	Cognitive domains within 7 days of Stroke Functional outcomes at 6 months post-stroke	Hierarchical multiple regression with the following variables Step 1: Age, sex, race, NIHSS, Education, stroke diagnosis, marital status Step 2: Berg balance scale, supine to sit, sit to stand, ambulation Step 3: SBT, TMT A & B, Cancellation, BNT
Laakso et al., 2019	Executive Function Subdomains: Stroop (response inhibition) TMT B and WCST (Set shifting) Verbal and design fluency (Initiation) WCST and Questioning task (strategy formation)	Activities of Daily Living (mRS)	Cognitive variable at 3 months post-stroke. Functional outcome at 15-months post-stroke	Logistic regression models. model 1 = adjusted for age, sex and years of education model 2 = additionally adjusted for NIHSS
Park et al., 2016	Executive Function (Verbal Fluency Test)	Activities of Daily Living (KBI and mRS)	Cognitive variables within 1 month of stroke.	Multiple linear regression examining relationship between cognitive domain and improvement in KBI scores between baseline and 3 months. Other variables included in regression were age, sex, hemiplegic side, initial K-MBI, FMSA.

	Visuospatial perception/construction (Construction Praxis test)		KBI at 3 months post-stroke. mRS at 6 months post-stroke.	Stepwise logistic regression carried out to examine relationship cognitive domain at baseline at mRS at follow-up. Other variables entered into the regression were age, sex, FMSA
Shea-Shumsky et al., 2019	Executive Function (Behavioural Dyscontrol Scale)	Activities of Daily Living (Basic ADL scale)	Cognitive variables at discharge. Functional outcomes at 3, 6, 9- and 12-months post discharge.	Stepwise regression analysis. Block 1 included age, sex, and education, Block 2 included the Charlson Comorbidity Index and pain severity, and Block 3 contained ADLs prior to admission. Block 4 included the Geriatric Depression Scale, MMSE, and BDS
Park et al., 2017	Executive Function (EF index of K-VCIH) Visuospatial Perception/Construction (ROCF) Visuomotor Speed (VCIHS digit symbol coding and TMT A)	Activities of Daily living (KBI and mRS)	Cognitive variables within 3 months of stroke (mean=20.2 *SD* =16.1 days) Functional outcome at discharge (mean length of stay =37, *SD* =22.7 days)	T-test using index scores and mRS and KBI, followed by univariate regression with individual subtest scores and KBI. Multivariate linear regression with all variables which were significant in Univariate analysis and KBI. Variables were: Age, length of stay, baseline scores on KBI, mRS, MMSE and NIHSS, ROCF, digit-symbol coding and other EF measures.
Nys et al., 2005	Executive Function (Brixton Spatial Anticipation Test, the Visual	Activities of Daily Living (BI)	Cognitive variables within 3 weeks of Stroke.	Univariate analysis looking at the association between cognitive domain scores at baseline and BI/FAI at follow-up.

	Elevator and letter fluency) Visuospatial perception/construction (Judgment of Line Orientation-SF, Test of Facial Recognition-SF and the ROCF– copy) Visual Memory (Corsi Block span, the ROCF-delay, the WMS–Visual Reproduction, and the Location Learning Task)	Instrumental Activities of Daily Living (FAI)	Functional outcomes between 6- and 10 months.	Two stepwise multiple logistic regressions for each outcome measure, one containing all cognitive variables which were significantly associated in univariate analyses and one containing both cognitive and medical predictor values which were significant in univariate analyses.
Cumming et al., 2014	Executive Function (TMT B, Controlled Oral Word Association Test and animal and letter fluency) Visuospatial Perception/Construction (ROCF-copy and WAIS-R block design)	Quality of Life (AQoL)	Cognitive domains at 3 months post-stroke. Functional outcomes at 6 months post-stroke	Univariate linear regression with Bonferroni adjusted significance. Multivariate regression with cognitive domains which were significant in univariate analysis. Controlling for age, stroke severity and depression
Nys et al., 2006	Executive Function (Brixton Spatial Anticipation Test, the Visual Elevator, letter fluency, the Stroop test, Semantic Fluency, and the Zoo test)	Quality of Life (SS-QoL)	Cognitive variables within 3 weeks post-stroke. Functional outcomes at 6-	Univariate regression. Multiple stepwise regression with the following variables entered: EF, visual memory, neglect, VSPC, age, education, lesion volume, hypercholesterolaemia, alcohol intake, cognitive impairment (MoCA), NIHSS, and BI.

	Visuospatial Perception/Construction (Judgment of Line Orientation, Test of Facial Recognition, the Rey-Osterrieth Complex Figure– copy, and the WAIS-III block design. Visual Memory (Corsi Block span, the ROCF-delay, the WMS–Visual Reproduction, and the Location Learning Task)		10 months post baseline.	
Hochstenbach et al., 2001	Executive Function (TMT A &B) Visuospatial Perception/Construction (WAIS block design, The Bobertag, a clock drawing task, the copying task from the BIT, and Money's Road Map) Visuomotor Speed (TMT A and WAIS digit-symbol coding)	Quality of Life (SIP)	Cognitive variables between 5- and 293-days post-stroke (mean: 72.2). Functional outcome at 9 months post-stroke	Univariate ANOVA and Chi Squared test. Forward-backwards stepwise regression with all variables which were significant in univariate analyses and the following participant variables: Educational level, gender, age, the side of the lesion (left, right), stroke type (hemorrhage, infarction), and location (subcortical, cortical).

Adamit et al., 2015	Executive Function (DEX and EFPT)	Participation in meaningful Activities (OQ)	Cognitive domain at 3 months Post-stroke. Functional outcome at 6 months post-stroke.	Stepwise regression with the following variables entered: Age, Gender, Education, NIHSS, MoCA score, EFPT, DEX and QoL recovery
Desrosiers et al., 2008	Executive Function (Stroop) Visuospatial Perception/Construction (VPT-Vertical)	Participation (LIFE-H; Divided into social roles and ADLs)	Cognitive domains measured at 2-3 weeks. Functional outcomes at 3 months and 6 months post-stroke	Stepwise multiple regression was run with multiple variables until the model of variables explaining most variance was found. This was performed separately for patients recruited from acute care and an inpatient rehabilitation facility.
Kapoor et al., 2019	Executive Function (TMT A & B) Visuomotor Speed (WAIS digit-symbol coding)	Participation (FAI) Reintegration (RNLI)	Cognitive variables at first stroke clinic visit. Functional outcomes between 1- and 2-years post-stroke.	Bivariate correlations between individual cognitive tests and outcome measure scores with alpha set at 0.05.
Ownsworth & Shum., 2008	Executive Function (Health and Safety sub-test, FAS Test, Five-Point test, Key Search Test and Tinkertoy Test)	Reintegration (SPRS) Employment Outcomes	Cognitive variables at enrollment study (mean = 2.1 years post-stroke).	Bivariate correlations to examine relationship between EF variables and reintegration scores. (For reintegration outcomes) *t*-test with covariate of time to look at ability of EF at baseline to discriminate between employment outcomes at follow-up. *t*-tests run with neglect patients included and excluded. (For employment outcomes)

			Functional outcome measures at 12-month follow-up.	
Narasimhalu et al., 2011	Visuospatial Perception/Construction (WMS- visual reproduction copy task, clock drawing, WAIS block design.) Visual Memory (picture recall and visual reproduction subtests of WMS) Visuomotor Speed (Digit symbol coding, digit modality test, digit cancellation, and maze task)	Activities of Daily Living (mRS)	Cognitive variables within 3-4 months of stroke. Functional outcomes measured at annual follow-up for up to 5 years.	Cox proportional hazards model regression Univariate and Multivariate. Multivariate regression was adjusted for stroke subtype, age, sex, diabetes mellitus, hypertension, as well as the treatment allocation
Wagle et al., 2011	Visuospatial Perception/Construction (RBANS VCI index) Visuomotor Speed (Coding subtest from the RBANS)	Activities of Daily Living (mRS)	Cognitive variables between 2-3 weeks post-stroke. Functional outcomes at 13-month follow-up.	An adjusted and unadjusted linear regression with RBANS index scores and an adjusted and unadjusted multiple linear regression with individual subtest scores.
Van Zandvoort et al., 2005	Visuospatial Perception/Construction	Activities of Daily Living (BI and mRS)	Cognitive variables measured	Bivariate correlations.

	(Judgement of Line Orientation Test, Test of Facial Perception-short form, ROCFT-copy) Visuomotor Speed (TMT-A)	Quality of Life (MOS-SF36 and VAS)	between 4- and 20-days post-stroke Functional variables measured between 12- and 24-months post-stroke	
Larson et al., 2003	Visuospatial Perception/Construction (RBANS VCI)	Activities of Daily Living (FIM; cognitive and motor) Instrumental Activities of Daily Living (FAI) Handicap (mCRAIG hospital assessment and reporting technique)	Cognitive variables measured during inpatient rehabilitation Functional measures 6 months post discharge.	Bivariate correlational analysis for RBANS index with each outcome measure. Bonferroni adjusted significance set at $p = .002$. Multiple linear regression with all noncollinear RBANS indices and each outcome measure.

Appendix D: Summary of sampling method and participant characteristics for studies included in review

Study	Country	Sample (size and method)	Sampling and Inclusion/exclusion criteria	Age/Sex	Diagnostic criteria and Stroke factors	Prevalence of domain specific cognitive impairments.
(Adamit et al., 2015)	Israel	Consecutive, opportunistic sampling from emergency department. 249 patients experiencing first time Stroke	Inclusion/exclusion criteria: Participants with mild stroke (NIHSS ≤ 5), above the age of 50, with no previous relevant neurological or psychiatric disease, who did not have dementia or aphasia.	Mean age: 68.6 Range: 50-92 Sex: Not reported.	First time acute ischemic stroke (as diagnosed using NICE criteria) Mean NIHSS scale score was 2.2 (SD = 1.5).	Not reported.
(Bertolin et al., 2018)	United States	Consecutive opportunistic sampling from the Brain Recovery Core (BRC; Lang et al., 2011) a collaborative endeavour among a university-affiliated medical centre, an acute care hospital, and a rehabilitation institute 498 Stroke patients	Inclusion criteria: Patient received clinical services for an acute stroke through the BRC between 2010 and 2014, the patient voluntarily provided informed consent, the patient completed the 6-month follow-up telephone interview. Exclusion criteria: A National Institute of Health Stroke Scale (NIHSS) score of >16 (indicating severe stroke), an NIHSS Aphasia item score of 2 or 3, indicating severe to global aphasia, a NIHSS Dysarthria item score of 2 or 3, indicating severe dysarthria or intubation.	Age: Mean= 64.50 ± 14.54 Range = 21–98 Sex: 51% male, 49% female	Stroke type: Ischemic = 60.60 (302) Haemorrhagic =10.80 (54) Unknown =28.50 (142)	Not reported.
(Cumming et al., 2014)	Australia	Consecutive opportunistic sample from Acute Stroke Unit 60 Stroke patients	Inclusion criteria: Confirmed stroke. Exclusion criteria:	Mean age: 72.1 (13.9) range: 30-95	Mild stroke: 44 (76%) Moderate stroke: 11 (19%)	Not reported.

			-<18 years old, Unconscious on admission, major visual, hearing or language impairment. -	Male: 73% Female: 27%	Severe stroke: 3 (5%) as measured by NIHSS Stroke type: TACI: 4 (7%) PACI: 21 (36%) POCI: 20 (35%) LACI: 8 (14%) ICH: 5 (19%) Left side lesion: 24 (41%) Right side: 34 (59%)	
(Desrosiers et al., 2008)	Canada	Opportunistic consecutive sample upon discharge from acute care hospital, inpatient rehabilitation unit and geriatric day hospital. 86 stroke patients from acute care group 111 stroke patients from rehabilitation centre	Inclusion/exclusion Criteria: Have suffered at least one stroke according to WHO criteria, over 65, being discharged from ACH, IRU or GDH, be returning to live at home, not have severe cognitive disorder (based on clinical judgement, defined as inability to follow simple discussion and lack of awareness of involvement in study).	Mean age AC: 77.2 (7.1) Mean age RC: 76.7 (7.0) Male AC: 48 (55.7%) Male RC: 53 (47.7%)	Stroke diagnosed according to WHO criteria. Type of stroke: Ischemic stroke AC: 77 (92.8%) RC: 95 (88.8%)	Not reported.
(Kapoor et al., 2019)	United Kingdom	Opportunistic consecutive sample of patients referred to a stroke prevention clinic. 124 stroke patients, a 'subset' of whom completed the neuropsychological tests. Article	Inclusion criteria: a definite diagnosis of stroke by a neurologist, initial clinic visit and screening within 6 months of the stroke and a follow-up assessment within predetermined follow-up window of 2.25 years (6 months) after initial clinic visit.	Mean age: 66.3 (15.7) Male: 65 (52.4%)	Stroke diagnosis confirmed by neurologist. Stroke Type: Ischemic: 116 (93.5%) Intracerebral Haemorrhage: 8 (6.5%)	Not reported.

		did not provide a numerical figure for this.	Exclusion criteria: Severe aphasia, severe motor dysfunction, and non-fluent English			
(Laakso et al., 2019)	Finland	Opportunistic consecutive sample of patients admitted to hospital emergency room 62 stroke patients	No inclusions/exclusion criteria provided.	Mean age: 68.4 (7.7) Males: 32 (51.6%)	NIHSS stroke symptom classification: No stroke symptoms: 21 (43.5%) Minor stroke: 26 (41.9%) Moderate Stroke: 8 (12.9%) Moderate to severe stroke: 1 (1.6%)	Not reported.
(Narasimhalu et al., 2011)	Singapore	Subset of participants recruited as part of a stroke prevention and treatment trial (ESPRIT). Opportunistic consecutive sample of patients with recent TIA or nondisabling ischemic stroke who were seen in the Singapore General Hospital 207 ischemic stroke patients	Inclusion criteria: Within 6 months of a transient ischemic attack (including transient monocular blindness) or a nondisabling ischemic stroke (grade ≤3 on the modified Rankin scale) of presumed arterial origin. Exclusion criteria: a possible cardiac source of embolism, high-grade carotid stenosis for which carotid endarterectomy or endovascular treatment was planned, moderate to severe leukoaraiosis on brain imaging (for randomization into anticoagulation), any blood coagulation disorder, any contraindication and dementia.	Mean age: 54 (10) Male: 163 (77%)	Stroke type: TIA: 59 (28%) POCI/LACI: 143 (67%) TACI/PACI: 10 (5%)	Not reported.
(Nys et al., 2006)	The Netherlands	Opportunistic consecutive sample of patients admitted to 3 stroke units. 91 stroke patients	Inclusion Criteria: Patients with a first-ever ischaemic stroke or primary intracerebral haemorrhage.	Mean age: 61.6 (13.2) Female: 48.4%	Diagnosis of stroke was based on both the presence of acute neurological symptoms	Not reported.

			Exclusion criteria: Pre-existing depression as diagnosed by general practitioner/psychiatrist, or history that might influence outcome, i.e. history of drug abuse, pre-existent dependence in activities of daily living, or pre-existent dementia (as defined by a score of 3.6 or higher on the short Informant Questionnaire on Cognitive Decline in the Elderly — IQCODE Dutch Version), Older than 85 years, and patients who could not be examined within the first 21 days post-stroke due to severe disturbances in consciousness or inability to comprehend task instructions.		and a compatible lesion on CT or MRI scan. Stroke severity: NIHSS > 7 75.8 Lesion side: Left lesion: 42.5 Right lesion: 43.7 Infratentorial: 13.8	
(Ownsworth & Shum, 2008)	Australia	Opportunistic consecutive sample from two hospitals and a community-based rehabilitation service. 27 stroke patients	Inclusion/exclusion criteria: A medical diagnosis of stroke, adequate communication skills to complete the assessments and had a relative who was available to participate in the study. An additional criterion was that individuals were employed or in tertiary studies at the time of their stroke.	Mean age: 47.3 (10.7) Range: 24 – 62 Male; 14 (52%)	Lesion side: Left 12 (44%) Right 11 (41%) Bilateral 4 (15%)	Not reported.
(J. Park et al., 2016)	Korea	Retrospective cohort sample from inpatient rehabilitation centre notes. 40 stroke patients	All stroke patients who underwent neuropsychological evaluation within 4 weeks were included. Exclusion criteria: Recurred stroke, previous history of psychiatric problems (e.g., depression, drug or alcohol abuse), previous history of traumatic brain injury or degenerative brain disease, other causes of disabilities (amputation, bedridden due to medical conditions, etc.), noncommunicable status (e.g., aphasia), and hemineglect.	Mean age: 68.05 (10.17) Sex: 50% Male	Stroke type: Haemorrhagic stroke: 7 Ischemic stroke: 29 Lesion side: Left: 21 Right:17 Bilateral: 2	Not reported.

(Shea-Shumsky et al., 2019)	United States	Opportunistic sample from Medicare beneficiaries in any of 14 skilled nursing facilities (SNFs) or rehabilitation units that participated in a study of rehabilitation outcomes across the United States. 246 stroke patients	Inclusion criteria: Inpatients receiving rehabilitation services after discharge from an acute-care hospitalization for stroke. Exclusion Criteria: Patients who did not speak English, were comatose on admission, or who were receiving hospice care.	Mean age: 77.5 (6.85) Age range: 65-97 Male: 96 (39%)	Stroke type: Brainstem: 17 (8.3%) Ischemic: 179 (87.3%) Haemorrhagic: 26 (12.7%) Intracerebral Haemorrhage: 17 (8.3%) Subarachnoid haemorrhage: 1 (0.5%) Intraventricular haemorrhage: 7 (3.4%) Subdural haemorrhage: 1 (0.5%). Lesion side: Left: 79 Right: 97 Bilateral: 29 Could not define 41	Not reported.
(Wagle et al., 2011)	Norway	Opportunistic consecutive sample from Stroke rehabilitation unit. 194 stroke patients.	Inclusion criteria: patients diagnosed with an ischaemic or haemorrhagic stroke. Exclusion criteria: nonfluency in the Norwegian language, severe visual or hearing impairment, alcohol or drug abuse, or previously being treated for a psychiatric illness	Mean age: 79.6 (10.7) Female: 75 (46%)	Stroke type: Ischemic stroke: 134 (82%) Haemorrhagic: 26 (16%) Haemorrhagic transformation: 3 (2%) OSC type: Taci: 36 (22)	Not reported.

					Paci: 64 (39) Laci: 50 (31) Poci: 13 (8) Severity: Mean NIHSS score: 7.8 (7.5)	
(S. H. Park et al., 2017)	Korea	Patients consecutively recruited from stroke rehabilitation centre. 104 stroke patients.	Inclusion criteria: Hospitalized within 7 days of symptom onset at the neurology or neurosurgery department, confirmed stroke by computed tomography or magnetic resonance images, transferred to the rehabilitation centre and completed the 60-minute K-VCIHS neuropsychological assessments within 3 months after stroke onset. Exclusion criteria: Patients who were transferred to other departments for concomitant events or patients who died during rehabilitation.	Mean age: 66.5 (11.7) Male: 76 Female: 28	Stroke Type: Ischemic: 82 Haemorrhagic: 22 Lesion Side: Right side: 54 Left side: 44 Both:6	Memory Impairment: 59.6% Visuospatial impairment: 62.5% Executive impairment: 60.5%
(G. M. S. Nys et al., 2005)	The Netherlands	Opportunistic consecutive sample of patients admitted with first ever stroke at 3 hospitals 168 stroke patients	Inclusion criteria: Confirmed either ischemic stroke or primary intracerebral haemorrhage. Exclusion criteria: pre-existing impairment or history that might influence cognitive or functional outcome, i.e., history of drug abuse, pre-existent dependence in activities of daily living, or pre-existent cognitive decline (as defined by a score of 3.6 or higher on the short Informant Questionnaire on Cognitive Decline in the Elderly–IQCODE Dutch version), age older than 85 years, and inability to be	Mean age: 60.1 SD 14.2 Male: 54.1%	Diagnosis of stroke was based on the presence of both an acute focal deficit and an associated lesion on CT or MRI. Stroke type: Ischemic: 90.1% Haemorrhagic: 9.9% Lesion location: Left: 43.8 %	Visual perception and Construction: 30.6% Executive functioning: 29.7% Visual memory: 16.2%

			examined within the first 21 days poststroke due to severe disturbances in consciousness or inability to comprehend task instructions.		Right: 41 % Brainstem/Cerebellum: 15.2% Stroke severity: Mean NIHSS score: 5 (0-18)	
(Larson et al., 2003)	USA	Consecutive Opportunistic sample of Stroke patients admitted to rehabilitation centre 158 Stroke patients	Inclusion/exclusion criteria: Onset of stroke within the preceding 6 months, no other neurological condition that could affect cognition, no substance abuse at the time of stroke, sufficient alertness to respond to questions (either verbally or through gesture), and use of at least one upper extremity.	Mean age: 65 Range: 31–85 Female: 59%	Lesion Side: Left: 32.4% Right: 61.7% Bilateral: 5.9%	Not reported.
(Van Zandvoort et al., 2005)	The Netherlands	Consecutive opportunistic sample from university medical Centre. 57 Stroke Patients	Inclusion/exclusion criteria: A first-ever symptomatic brain infarct, a maximum age of 80 years, Modified-Rankin Scale score in the range from 2 to 4, command of the Dutch language, a stay of more than 1 day at the stroke unit, independent living in the community prior to the stroke, and no psychiatric history or comorbidity that could be of influence on cognitive functioning.	Mean age: 56 (S.D. = 16) Range: 19–80) 31 males, 26 females	Stroke type: All Ischemic Lesion side: Left side: 21 Right side: 27 Bilateral: 4 Infratentorial: 5	Digit Span (WAIS): 35% Corsi Block-Tapping Task: 8% CFT – delayed recall: 77% TMT (Part A1): 65% TMT (A2): 71% TMT (B): 71%

						CFT – copy: 54% Test of Facial Perception: 29% Judgement of Line Orientation: 68%
(Hochsten bach et al., 2001)	Denmark	Consecutive Opportunistic sample from several hospitals. 164 Stroke patients	Inclusion criteria: Confirmed Stroke, less than 70 years old. Exclusion criteria: Any other major physical diseases or mental disorders, or if a reliable assessment of neuropsychologic functions was not possible (e.g., in patients with a persistent impairment in consciousness), Aphasic patients were excluded only if they were unable to respond adequately to "yes" or "no" questions. At follow-up, an additional inclusion criterion was that patients had to be living at home, not in an institution, to ascertain the most optimal participation in life.	Mean age: 55.1 ± 10.9 Male: 102 Female: 62	Stroke diagnosed by topography. Stroke type: Ischemic Stroke: 142 Haemorrhagic Stroke: 22 Lesion Side: Left side: 81 Right Side: 75 Bilateral: 6	Not reported.

Appendix E: Summary of quality appraisal scores using Down and Black's quality appraisal checklist

Citation	Reporting							External Validity			Internal Validity				Power	Overall
	Hypothesis	Outcome	Sample	Findings	Variance	P Values	Response Rate	Those asked	Those agreed	Location	Data-Dredging	Statistical tests	Outcome Measures	Confounds	Power Calculation	
(Nys et al., 2005)	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	N	12
(Van Zandvoort et al., 2005)	Y	Y	N	N	Y	N	N	N	N	N	N	Y	N	N	N	4
(Adamit et al., 2015)	Y	Y	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	N	11
(Bertolin et al., 2018)	Y	Y	Y	Y	Y	N	N	N	N	Y	Y	Y	Y	Y	N	10
(Cumming et al., 2014)	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	Y	N	N	8
(Desrosiers et al., 2008)	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	N	N	11
(Hochstenbach et al., 2001)	Y	Y	Y	Y	Y	Y	N	N	N	N	Y	Y	Y	Y	N	10
(Kapoor et al., 2019)	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	Y	N	N	10
(Laakso et al., 2019)	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	N	12
(Larson et al., 2003)	Y	Y	Y	N	N	N	N	N	N	Y	Y	Y	Y	N	N	7

(Narasimhalu et al., 2011)	**Y**	**Y**	**N**	**N**	**Y**	**Y**	**Y**	**N**	**N**	**N**	**Y**	**Y**	**Y**	**N**	**N**	**9**
(Nys et al., 2006)	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**12**
(Ownsworth & Shum, 2008)	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**N**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**10**
(Park et al., 2016)	**Y**	**Y**	**N**	**Y**	**Y**	**Y**	**N**	**N**	**N**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**10**
(Park et al., 2017)	**Y**	**Y**	**N**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**11**
(Shea-Shumsky et al., 2019)	**Y**	**Y**	**N**	**N**	**Y**	**Y**	**N**	**N**	**N**	**N**	**Y**	**Y**	**Y**	**Y**	**N**	**8**
(Wagle et al., 2011)	**Y**	**Y**	**Y**	**N**	**Y**	**Y**	**Y**	**N**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**12**

Appendix F: Outline of SPANS subtests and indices (extracted from the SPANS manual; Burgess, 2014)

The SPANS contains 30 subtests that make-up the following 7 indices.

Orientation Index (ORI)

The ORI measure's orientation to person, time, place, condition and political leadership. Scores are graded according to exact, near or distal answers. For example, an answer 1 year younger than their age is scored higher than 20 years under which suggests different degrees of intact mental status. All answers are easily verifiable. The following subtests are included:

- Orientation
    - Person (name, date of birth, age)
    - Time (of day, day of week, month, year)
    - Place (city and type/name of place)
    - Condition (awareness of condition)
    - Political Leadership (present and past)
- Time estimation (how long was duration of testing)

Attention/Concentration Index (ACI)

The ACI measure's aspects of attention and concentration including span/capacity, sustained and divided attention, response inhibition, mental control (counting backwards) and mental monetary calculations. Counting backwards and mental calculations tasks also include speed of information components which can earn extra points. The following subtests are included:

- Digit Span Forward
- Digit Span Backward
- Sustained and Divided Listening – Round 1
- Sustained and Divided Listening – Round 2
- Counting Backwards
- Monetary Calculations

Language Index (LAI)

The LAI incorporates measures to screen for aphasia, alexia and agraphia or detect other language disturbances. Subtests include confrontation naming (with a scoring scheme that reflects any need for phonetic cues), repetition, comprehension and free expressive language/verbal reasoning. Fluency of non-fluency or non-fluency of speech is evaluated through observation. Brief screenings of reading (reading and following written commands) and writing (original and dictated sentences) are also included, providing an ecologically valid assessment of everyday activities. The following subtests are included:

- Repetition
- Naming
- Yes/No Questions

- Following Directions
- Reading
- Writing Sentences
- Similarities

Memory and Learning Index (MLI)

The MLI measures memory and learning via several means, including for verbal and visual material. Two 'recall' subtests are composed of 'learning' trials (i.e. repeated lists and associative learning), and two are composed of only a single exposure to the material. Three of the subtests contain a five-minute delay of an intervening (but unrelated and non-confounding) activity before the recall trial. Ecological validity was included as a feature (i.e. learning a shopping list). It includes the following subtests:

- Object Recall
- Figures Recall
- List Learning
- List Recall
- List Recognition
- Symbol-Word Paired-Associates

Visuo-Motor Performance Index (VPI)

The VPI measures various visuo-spatial/visuo-perceptual and motor capabilities, including screening for spatial impairment and/or object perceptual agnosia. Following a visual screening test, the index includes visual attention and visual recognition memory, copy of geometric figures, visuo-motor coding, spatial and object perception, reading emotion in facial expressions and visual concepts, with 'free-choice' and 'recognition' scoring criteria. It includes the following subtests:

- Object Recognition
- Spatial Decision
- Unusual Views
- Figures Copy
- Letter-Number Coding
- Figures Recognition
- Facial Expressions
- 3-and-1 Concept Test

Efficiency index (ECI)

The ECI combines the subtests with a timed element, thus evaluating the speed of reacting, thinking, scanning, and visuo-motor movement in unison, for an overall estimate of the efficiency of processing. It includes the following subtests:

- Sustained and Divided Listening -Round 2
- Spatial Decision
- Letter-Number Coding

- Counting Backwards
- Monetary Calculations

Conceptual Flexibility Index

The CFI combines two subtests that each possess elements of concept formation, thinking laterally and flexibility, and combining concepts into a superordinate category, with both visual and verbal elements. It includes the following subtests:

- Similarities
- 3-and-1 Concept Test

Errors for sustained attention commissions, perceptual naming errors and memory intrusions are also scored.

Appendix G: Output for *t*-test and McNemar's test to assess differences between sample demographics

Paired Samples t-Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Age - Age	.173	3.161	.310	-.442	.788	.558	103	.578

McNemar's clinical & control

	control	
clinical	Male	female
Male	42	44
female	6	12

McNemar's Test Statistics[a]

	clinical & control
N	104
Chi-Square[b]	27.380
Asymp. Sig.	.000

a. McNemar Test

b. Continuity Corrected

Appendix H: Two-way repeated measures ANOVA to assess impact of difference in sex ratios between samples on SPANS scores

SPANS total score Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
sample	Sphericity Assumed	26661.253	1	26661.253	54.938	.000
	Greenhouse-Geisser	26661.253	1.000	26661.253	54.938	.000
	Huynh-Feldt	26661.253	1.000	26661.253	54.938	.000
	Lower-bound	26661.253	1.000	26661.253	54.938	.000
Error(sample)	Sphericity Assumed	8250.059	17	485.298		
	Greenhouse-Geisser	8250.059	17.000	485.298		
	Huynh-Feldt	8250.059	17.000	485.298		
	Lower-bound	8250.059	17.000	485.298		
sex	Sphericity Assumed	222.253	1	222.253	.381	.546
	Greenhouse-Geisser	222.253	1.000	222.253	.381	.546
	Huynh-Feldt	222.253	1.000	222.253	.381	.546
	Lower-bound	222.253	1.000	222.253	.381	.546
Error(sex)	Sphericity Assumed	9929.559	17	584.092		
	Greenhouse-Geisser	9929.559	17.000	584.092		
	Huynh-Feldt	9929.559	17.000	584.092		
	Lower-bound	9929.559	17.000	584.092		
sample * sex	Sphericity Assumed	408.503	1	408.503	.859	.367
	Greenhouse-Geisser	408.503	1.000	408.503	.859	.367
	Huynh-Feldt	408.503	1.000	408.503	.859	.367
	Lower-bound	408.503	1.000	408.503	.859	.367
Error(sample*sex)	Sphericity Assumed	8080.559	17	475.327		

	Greenhouse-Geisser	8080.559	17.000	475.327		
	Huynh-Feldt	8080.559	17.000	475.327		
	Lower-bound	8080.559	17.000	475.327		

ORI Index Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
sample	Sphericity Assumed	276.125	1	276.125	22.300	.000
	Greenhouse-Geisser	276.125	1.000	276.125	22.300	.000
	Huynh-Feldt	276.125	1.000	276.125	22.300	.000
	Lower-bound	276.125	1.000	276.125	22.300	.000
Error(sample)	Sphericity Assumed	210.500	17	12.382		
	Greenhouse-Geisser	210.500	17.000	12.382		
	Huynh-Feldt	210.500	17.000	12.382		
	Lower-bound	210.500	17.000	12.382		
sex	Sphericity Assumed	.125	1	.125	.008	.929
	Greenhouse-Geisser	.125	1.000	.125	.008	.929
	Huynh-Feldt	.125	1.000	.125	.008	.929
	Lower-bound	.125	1.000	.125	.008	.929
Error(sex)	Sphericity Assumed	257.250	17	15.132		
	Greenhouse-Geisser	257.250	17.000	15.132		
	Huynh-Feldt	257.250	17.000	15.132		
	Lower-bound	257.250	17.000	15.132		
sample * sex	Sphericity Assumed	.222	1	.222	.014	.906

	Greenhouse-Geisser	.222	1.000	.222	.014	.906
	Huynh-Feldt	.222	1.000	.222	.014	.906
	Lower-bound	.222	1.000	.222	.014	.906
Error(sample*sex)	Sphericity Assumed	264.903	17	15.583		
	Greenhouse-Geisser	264.903	17.000	15.583		
	Huynh-Feldt	264.903	17.000	15.583		
	Lower-bound	264.903	17.000	15.583		

MLI Index Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
sample	Sphericity Assumed	3770.014	1	3770.014	41.396	.000
	Greenhouse-Geisser	3770.014	1.000	3770.014	41.396	.000
	Huynh-Feldt	3770.014	1.000	3770.014	41.396	.000
	Lower-bound	3770.014	1.000	3770.014	41.396	.000
Error(sample)	Sphericity Assumed	1548.236	17	91.073		
	Greenhouse-Geisser	1548.236	17.000	91.073		
	Huynh-Feldt	1548.236	17.000	91.073		
	Lower-bound	1548.236	17.000	91.073		
sex	Sphericity Assumed	.347	1	.347	.002	.963
	Greenhouse-Geisser	.347	1.000	.347	.002	.963
	Huynh-Feldt	.347	1.000	.347	.002	.963
	Lower-bound	.347	1.000	.347	.002	.963
Error(sex)	Sphericity Assumed	2657.903	17	156.347		
	Greenhouse-Geisser	2657.903	17.000	156.347		

	Huynh-Feldt	2657.903	17.000	156.347		
	Lower-bound	2657.903	17.000	156.347		
sample * sex	Sphericity Assumed	62.347	1	62.347	.598	.450
	Greenhouse-Geisser	62.347	1.000	62.347	.598	.450
	Huynh-Feldt	62.347	1.000	62.347	.598	.450
	Lower-bound	62.347	1.000	62.347	.598	.450
Error(sample*sex)	Sphericity Assumed	1771.903	17	104.230		
	Greenhouse-Geisser	1771.903	17.000	104.230		
	Huynh-Feldt	1771.903	17.000	104.230		
	Lower-bound	1771.903	17.000	104.230		

LAI Index Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
sample	Sphericity Assumed	953.389	1	953.389	22.586	.000
	Greenhouse-Geisser	953.389	1.000	953.389	22.586	.000
	Huynh-Feldt	953.389	1.000	953.389	22.586	.000
	Lower-bound	953.389	1.000	953.389	22.586	.000
Error(sample)	Sphericity Assumed	717.611	17	42.212		
	Greenhouse-Geisser	717.611	17.000	42.212		
	Huynh-Feldt	717.611	17.000	42.212		
	Lower-bound	717.611	17.000	42.212		
sex	Sphericity Assumed	220.500	1	220.500	3.603	.575
	Greenhouse-Geisser	220.500	1.000	220.500	3.603	.575
	Huynh-Feldt	220.500	1.000	220.500	3.603	.575

	Lower-bound	220.500	1.000	220.500	3.603	.575
Error(sex)	Sphericity Assumed	1040.500	17	61.206		
	Greenhouse-Geisser	1040.500	17.000	61.206		
	Huynh-Feldt	1040.500	17.000	61.206		
	Lower-bound	1040.500	17.000	61.206		
sample * sex	Sphericity Assumed	272.222	1	272.222	5.383	.730
	Greenhouse-Geisser	272.222	1.000	272.222	5.383	.730
	Huynh-Feldt	272.222	1.000	272.222	5.383	.730
	Lower-bound	272.222	1.000	272.222	5.383	.730
Error(sample*sex)	Sphericity Assumed	859.778	17	50.575		
	Greenhouse-Geisser	859.778	17.000	50.575		
	Huynh-Feldt	859.778	17.000	50.575		
	Lower-bound	859.778	17.000	50.575		

ECI Index Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
sample	Sphericity Assumed	953.389	1	953.389	22.586	.000
	Greenhouse-Geisser	953.389	1.000	953.389	22.586	.000
	Huynh-Feldt	953.389	1.000	953.389	22.586	.000
	Lower-bound	953.389	1.000	953.389	22.586	.000
Error(sample)	Sphericity Assumed	717.611	17	42.212		
	Greenhouse-Geisser	717.611	17.000	42.212		
	Huynh-Feldt	717.611	17.000	42.212		
	Lower-bound	717.611	17.000	42.212		

sex	Sphericity Assumed	220.500	1	220.500	3.603	.075
	Greenhouse-Geisser	220.500	1.000	220.500	3.603	.075
	Huynh-Feldt	220.500	1.000	220.500	3.603	.075
	Lower-bound	220.500	1.000	220.500	3.603	.075
Error(sex)	Sphericity Assumed	1040.500	17	61.206		
	Greenhouse-Geisser	1040.500	17.000	61.206		
	Huynh-Feldt	1040.500	17.000	61.206		
	Lower-bound	1040.500	17.000	61.206		
sample * sex	Sphericity Assumed	272.222	1	272.222	5.383	.033
	Greenhouse-Geisser	272.222	1.000	272.222	5.383	.033
	Huynh-Feldt	272.222	1.000	272.222	5.383	.033
	Lower-bound	272.222	1.000	272.222	5.383	.033
Error(sample*sex)	Sphericity Assumed	859.778	17	50.575		
	Greenhouse-Geisser	859.778	17.000	50.575		
	Huynh-Feldt	859.778	17.000	50.575		
	Lower-bound	859.778	17.000	50.575		

CFI IndexTests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
sample	Sphericity Assumed	355.556	1	355.556	20.919	.000
	Greenhouse-Geisser	355.556	1.000	355.556	20.919	.000
	Huynh-Feldt	355.556	1.000	355.556	20.919	.000
	Lower-bound	355.556	1.000	355.556	20.919	.000
Error(sample)	Sphericity Assumed	288.944	17	16.997		

	Greenhouse-Geisser	288.944	17.000	16.997		
	Huynh-Feldt	288.944	17.000	16.997		
	Lower-bound	288.944	17.000	16.997		
sex	Sphericity Assumed	26.889	1	26.889	1.695	.210
	Greenhouse-Geisser	26.889	1.000	26.889	1.695	.210
	Huynh-Feldt	26.889	1.000	26.889	1.695	.210
	Lower-bound	26.889	1.000	26.889	1.695	.210
Error(sex)	Sphericity Assumed	269.611	17	15.859		
	Greenhouse-Geisser	269.611	17.000	15.859		
	Huynh-Feldt	269.611	17.000	15.859		
	Lower-bound	269.611	17.000	15.859		
sample * sex	Sphericity Assumed	37.556	1	37.556	2.916	.106
	Greenhouse-Geisser	37.556	1.000	37.556	2.916	.106
	Huynh-Feldt	37.556	1.000	37.556	2.916	.106
	Lower-bound	37.556	1.000	37.556	2.916	.106
Error(sample*sex)	Sphericity Assumed	218.944	17	12.879		
	Greenhouse-Geisser	218.944	17.000	12.879		
	Huynh-Feldt	218.944	17.000	12.879		
	Lower-bound	218.944	17.000	12.879		

ACI Index Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
sample	Sphericity Assumed	1192.347	1	1192.347	45.663	.000

	Greenhouse-Geisser	1192.347	1.000	1192.347	45.663	.000
	Huynh-Feldt	1192.347	1.000	1192.347	45.663	.000
	Lower-bound	1192.347	1.000	1192.347	45.663	.000
Error(sample)	Sphericity Assumed	443.903	17	26.112		
	Greenhouse-Geisser	443.903	17.000	26.112		
	Huynh-Feldt	443.903	17.000	26.112		
	Lower-bound	443.903	17.000	26.112		
sex	Sphericity Assumed	36.125	1	36.125	.653	.430
	Greenhouse-Geisser	36.125	1.000	36.125	.653	.430
	Huynh-Feldt	36.125	1.000	36.125	.653	.430
	Lower-bound	36.125	1.000	36.125	.653	.430
Error(sex)	Sphericity Assumed	940.125	17	55.301		
	Greenhouse-Geisser	940.125	17.000	55.301		
	Huynh-Feldt	940.125	17.000	55.301		
	Lower-bound	940.125	17.000	55.301		
sample * sex	Sphericity Assumed	70.014	1	70.014	1.402	.253
	Greenhouse-Geisser	70.014	1.000	70.014	1.402	.253
	Huynh-Feldt	70.014	1.000	70.014	1.402	.253
	Lower-bound	70.014	1.000	70.014	1.402	.253
Error(sample*sex)	Sphericity Assumed	849.236	17	49.955		
	Greenhouse-Geisser	849.236	17.000	49.955		
	Huynh-Feldt	849.236	17.000	49.955		
	Lower-bound	849.236	17.000	49.955		

Appendix I: Copy of REC ethical approval

The REC ethical approval letter provided to Dr Gerald Burgess for the original collection of data for the validation of the SPANS cannot be accessed at this point in time due to Covid-19 University restrictions. As soon as restrictions are lifted a copy will be attached.

Appendix J: Statement of epistemological position

For both the literature review and the experimental research report a positivist epistemological position was taken. Conclusions were derived using the scientific method, with measurable data collected and analysed using objective statistical techniques. The researcher perceived themselves as separate from the data for both projects as data were collected independently and anonymised before being given to them for analysis. The researcher has not carried out research in this area previously to this project, though they did have some experience on a neuropsychological placement making the concepts more familiar to them. The researcher's supervisor was the author of the SPANS assessment tool which was the focus of the experimental research project. The researcher tried to maintain an objective viewpoint and was aware of the potential for bias when interpreting statistical results.

Appendix K: Chronology of the research process

Research Activity	Dates
Experimental thesis-Original project	
Experimental project design	June-September 2018
First draft of research proposal	November 2018
Peer review of research proposal	April 2019
Preparation of proposal and documents for NHS ethics	May-August 2019
Sponsorship approval received	September 2019
Submission to NHS R&I and REC	October 2019
HRA and REC favourable opinion given	December 2019
R&I approval and confirmation of capacity	March 2020
Sponsor green light received	March 2020

Due to the situation with covid-19 data collection was not able to go ahead for this project. The project was abandoned at this point and new one started.

Experimental thesis-New project	
Planning for new project	April/May 2020
Receival of retrospective data for new project	June 2020
Thesis preparation	June-August 2020
Thesis Submission	15th August 2020
Literature review	
First draft	April 2020
Second draft	August 2020
Submission	15th August 2020

Appendix L: Confidentiality checklist

	Checked in Abstract	Checked in main text	Checked in appendices
Pseudonym or false initials used	N/A	N/A	N/A
Reference to pseudonym/false initials as a footnote	N/A	N/A	N/A
Removed any reference to names of Trusts/hospitals/clinics/services (including letterhead if including letters in appendices)	✓	✓	✓
Removed any reference to names/specific dates of birth/specific date of clinical appointments/addresses/ location of client(s), participant(s), relatives, caregivers, and supervisor(s).	✓	✓	✓
Removed/altered references to client(s) jobs/professions/nationality where this may potentially identify them.	✓	✓	✓
Removed any information that may identify the Trainee (consult with course staff if this will detract from the points the Trainee is making)	N/A	N/A	N/A
No Tippex or other method has been used to obliterate the original text – unless the paper is subsequently photocopied, and the Trainee has ensured that the obliterated text cannot be read	✓	✓	✓
The "find and replace" function in word processing has been used to check the assignment for use of client(s) names/other confidential information	✓	✓	✓