VOLUME I: RESEARCH COMPONENT

Validation of the Delis-Kaplan Executive Function System (D-KEFS) in participants with Traumatic Brain Injury

by

Yin-Ming Chan

A THESIS SUBMITTED TO THE UNIVERSITY OF BIRMINGHAM FOR THE DEGREE

OF DOCTORATE IN CLINICAL PSYCHOLOGY

Department of Clinical Psychology

School of Psychology

The University of Birmingham

April 2019

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

THESIS OVERVIEW

The thesis is submitted to the University of Birmingham in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology. This thesis is organised as two separate volumes.

Volume one comprises a systematic review and an empirical research study. The systematic review aims to summarise and critically evaluate the evidence about the clinical usefulness of the Delis-Kaplan Executive Function System (D-KEFS) with regard to the identification of the executive impairment between people suffering from acquired neurological pathologies and healthy individuals or between different clinical groups. A systematic search of literature databases identified 32 relevant journal articles. Studies reporting group mean comparisons in D-KEFS performance, neuroanatomical correlates of the D-KEFS and the diagnostic accuracy of the D-KEFS were eligible.

The empirical research study examines the validity of the D-KEFS in the evaluation of executive functioning in a sample of participants with Traumatic Brain Injury (TBI), using orthopaedic patients as study controls. To maximise reliability, D-KEFS Executive Functioning Indices (EFIs) were constructed as suggested by literature. The utility of the individual D-KEFS subtests and the constructed indices to TBI in terms of their ability to the detection of head injury was determined.

A public domain briefing document is also included in this volume, providing information about the systematic review and empirical study to a wider audience in an accessible manner.

The second volume consists of five Clinical Practice Reports (CPRs), completed during placements in an Improving Access to Psychological Therapy (IAPT) service, an older adults

community mental health team, a child and adolescent mental health team and an outpatient neurorehabilitation team. CPR 1 reports cognitive-behavioural and psychodynamic formulations of an adult male experiencing depression and anxiety. CPR 2 presents a service evaluation of a newly launched Long Term Condition (LTC) drop-in service within a primary mental health setting. CPR 3 is a single-case experimental design study, investigating the effectiveness of a cognitive-behavioural intervention for an old man with fear of falling. CPR 4 describes clinical work from an attachment perspective with a boy displaying anger and aggressiveness towards his mother. CPR 5 is an abstract of an oral presentation that outlines an assessment, formulation and intervention of a man diagnosed with Multiple Sclerosis (MS) who presents with apathy and social isolation.

Dedicated

To my father and friends

for all your encouragement, support and prayers

ACKNOWLEDGEMENTS

CONTENTS OF VOLUME I: RESEARCH COMPONENT

SYSTEMATIC REVIEW: THE CLINCIAL USEFULNESS OF DELIS-KAPLAN EXECUTIVE FUNCTION SYSTEM (D-KEFS) IN THE EVALUATION OF EXECUTIVE FUNCTIONS IN CLINICAL POPULATIONS WITH ACQUIRED NEUROLOGICAL CONDITONS

EMPIRICAL PAPER: INVESTIGATING THE VALIDITY OF THE DELIS-KAPLAN EXECUTIVE FUNCTION SYSTEM (D-KEFS) AS A NEUROPSYCHOLOGICAL ASSESSMENT TOOL FOR EXECUTIVE FUNCTIONS IN THE TRAUMATIC BRAIN INJURY (TBI) IN THE UK

PUBLIC DOMAIN BRIEFING PAPER

SYSTEMATIC REVIEW

EMPIRICAL PAPER

LIST OF APPENDICES FOR VOLUME I

CONTENTS OF VOLUME II: CLINICAL COMPONENT

CPR 1: Cognitive-behavioural and Psychodynamic formulations of David, a 34-year-old man experiencing depression and anxiety

CPR 2: Service evaluation of a newly launched Long Term Condition (LTC) drop-in service within a primary mental health setting

CPR 3: Investigating the effectiveness of a cognitive-behavioural intervention for John, a 74-year-old man with fear of falling using a single case experimental design

CPR 4: A case study of attachment-based treatment intervention for David, an 11-year-old boy and his mother

CPR 5: Oral presentation of an assessment, formulation and intervention for George, a 48-year-old man diagnosed with Multiple Sclerosis (MS) who presents with apathy and social isolation

LIST OF APPENDICES FOR VOLUME II

LIST OF ILLUSTRATIONS

LIST OF TABLES

SYSTEMATIC REVIEW:

THE CLINICAL USEFULNESS OF DELIS-KAPLAN EXECUTIVE FUNCTION SYSTEM (D-KEFS) IN THE EVALUATION OF EXECUTIVE FUNCTIONS IN CLINICAL POPULATIONS WITH ACQUIRED NEUROLOGICAL CONDITONS

by

Yin-Ming Chan

ABSTRACT

Background

Deficits in executive functioning are commonly reported in people with acquired neurological conditions. A comprehensive assessment of executive functioning is therefore particularly important for these population groups from a perspective of clinical practice. The Delis-Kaplan Executive Function System (D-KEFS) is a set of standardised tests that comprehensively assess executive functions in both children and adults. Although there is some evidence in support of the effectiveness of the D-KEFS, its clinical usefulness of identifying executive dysfunctions in people with acquired neurological conditions has not been systematically reviewed. The aim of this review was to summarise and critically evaluate the evidence about the clinical usefulness of the D-KEFS with regard to the identification of the impairment in executive functioning between people suffering from acquired neurological pathologies and healthy controls or between different clinical groups. Studies reporting group mean comparisons in D-KEFS performance, neuroanatomical correlates of the D-KEFS and the diagnostic accuracy of the D-KEFS were eligible.

Search methods

A systematic literature search in three databases (PsycINFO, MEDLINE(R) and EMBASE) was conducted. Search terms related to executive function, acquired neurological disorders, D-KEFS and clinical usefulness were combined to locate studies. The search was limited to articles published in peer-reviewed journals and in the English language, between 2001 and 2018. Both titles and abstracts were examined and reference lists of the included studies were also reviewed to identify additional papers.

Results

A total of thirty-two studies were finally included comprised the following: five studies examined the executive functioning of individuals with Traumatic Brain Injury (TBI); six focused on patients with focal brain lesions; thirteen conducted analysis on different types of neurodegenerative disorders; and eight focused on epilepsy. The selection of the D-KEFS subtests also varied across the reviewed studies.

Conclusions

The D-KEFS appears to be a useful evaluation tool of executive functioning, based on the available evidence. The findings indicated that participants with various acquired neurological conditions showed significant executive impairment, including committing more errors than healthy individuals. The performance on the D-KEFS was also correlated with frontal brain regions and other related brain circuitry. Moreover, the D-KEFS may have some value in discriminating people with neurological pathologies from healthy population, although the evidence available is insufficient. More research may need to be conducted in future.

Keywords

Executive function, executive dysfunction, Delis-Kaplan Executive Function System, D-KEFS, clinical usefulness, acquired neurological disorders

INTRODUCTION

What is executive functioning?

The evaluation of executive functioning is an essential component in neuropsychological assessment. Executive functions refer to a wide range of higher-order cognitive processes that are necessary for formulating goals, prioritising, organising and carrying out plans to complete tasks effectively (Cummings & Miller, 2007; Jurado & Rosselli, 2007). One definition of executive functions given by Crawford is as follows:

> The term 'executive functions' is a convenient shorthand for a set of behavioural competencies which include planning, sequencing, the ability to sustain attention, resistance to interference, utilisation of feedback, the ability to co-ordinate simultaneous activity, cognitive flexibility [...], more generally, the ability to deal with novelty. (Crawford, 1998, p. 209)

In instances of deficits in these mental capacities, a person's ability to generate effective goals and identify potential solutions to problems can be greatly compromised and this often results in a profound negative impact on many aspects of the individual's everyday life (Jurado & Rosselli, 2007). In other words, executive functions lie at the heart of initiating socially productive, independent, purposeful and goal-directed behaviours (Lezak, Howieson, Bigler & Tranel, 2012). Given the conceptualisation that executive functions involve the co-ordination of various complex cognitive processes in order to achieve a particular goal, this reflects the construct of executive functioning being multifaceted in nature rather than being a unitary concept (Struss & Alexander, 2000).

In fact, studying executive functions is criticised as difficult and challenging (Miyake, Emerson & Friedman, 2000). Factor analytic studies, however, do offer some support for a multifaceted perspective. Miyake et al. (2000), for example, postulated three aspects of executive functions (updating information in working memory, inhibiting responses and switching between tasks). Nevertheless, although these studies support the heterogeneous nature of executive functioning, relying on factor analysis to determine the constructs of executive functions can still be problematic (Suchy, 2016). Some researchers, for instance, have found six underlying dissociable factors from a battery of 19 executive function measures in a sample of 200 healthy individuals using exploratory factor analysis (Testa, Bennett & Ponsford, 2012). The six factors identified were: prospective working memory, task analysis, set-shifting and interference management, strategy generation and regulation, self-monitoring and self-maintenance, and response inhibition. Thus, it can be readily seen that the consensus on the number of components identified, or what underlying cognitive constructs are represented has still not been reached among researchers.

Challenges in the assessment of executive functions in neurologically impaired populations

The term "Acquired Neurological Conditions" encompasses a wide range of neurological abnormalities that develop after we are born. These pathologies include, but are not limited to, acquired brain insults (e.g., traumatic head injury), medical conditions-associated brain damage (e.g., cerebral vascular accident, epilepsy, or HIV infection), and neurodegenerative disorders such as dementias of different kinds, Parkinson's disease, or multiple sclerosis. Neuropsychological evidence indicates that individuals with these neurological conditions all exhibit deficits in executive functioning (Suchy, 2016). A comprehensive assessment of executive functioning is therefore particularly important for these population groups from a perspective of clinical practice. Not only does this help in the identification of the nature and

severity of executive dysfunctions, it is also useful in the monitoring of treatment response and the planning of rehabilitation strategies (Cicerone, 2005; Miyake, Emerson & Friedman, 2000). These kinds of information are valuable not just for the diagnosis of disorder but also for the evaluation of the progression of these brain diseases over time (Dubois, Slachevsky, Litvan & Pillon, 2000). For these reasons, a great deal of attention has been given to the development of reliable measures of executive functioning.

As critical as it is for the management of neurologically impaired individuals, a proper evaluation of executive functioning for this population group can be complicated. Firstly, the definitions of executive functions differ widely and there is a lack of general agreement regarding the construct (Miyake, Emerson & Friedman, 2000). This has in turn led to the heterogeneity of executive function tests in terms of their formats as well as of the number and types of executive processes assessed (Alvarez & Emory, 2006). In addition, the impairment of executive functioning may vary considerably due to the range of neural systems and cognitive mechanisms that are involved, as well as the variations in the pathological characteristics of the brain injury. Therefore, while exploring the extent to which a test instrument can sensitively distinguish various neurological pathologies from normal performance may present considerable challenges, it could be of immense interest to clinical practitioners.

Delis-Kaplan Executive Function System (D-KEFS)

Although there are a number of neuropsychological tests developed to assess executive functioning, several studies have suggested that many of these tests may not be sufficiently sensitive to detect executive dysfunctions in different clinical groups (Chan, Shum, Toulopoulou & Chen, 2008). Recently, a neuropsychological battery of tests, namely the Delis-Kaplan Executive Function System (D-KEFS), has been developed to comprehensively

evaluate a wide range of executive functioning in both verbal and non-verbal modalities for individuals from 8 to 89 years of age. The D-KEFS possesses its uniqueness over other traditional executive functioning tests because of its large representative standardisation sample and the addition of a qualitative error measurement (Goldberg & Bougakov, 2005; Delis, Kaplan & Kramer, 2001a). This standardised set of tests is made up of nine stand-alone subtests mostly derived from existing neuropsychological measures but many have been slightly modified in order to highlight the measurement of executive functioning (Swanson, 2005). Furthermore, each D-KEFS subtest can either be used individually as a stand-alone instrument or administered in combination with other D-KEFS subtests to provide a comprehensive tool for assessing a wide array of cognitive domains, including cognitive shifting, inhibition, problem solving, planning, creativity, reasoning, abstract thinking and concept formation (Strauss, Sherman & Spreen, 2006). Detailed description of the nine subtests is presented in Table 1:

Table 1: The nine D-KEFS subtests and associated domains assessed (Information from Strauss, Sherman & Spreen, 2006)

D-KEFS Subtests	Details	Key Areas Assessed
Trail Making Test (TMT)	This test comprises 5 conditions, the Visual Scanning, the Number Sequencing, the Letter Sequencing, the Number-Letter Switching and the Motor Speed. The examinee is asked to scan letters and numbers and mark the number 3 on condition 1; connect just the numbers in ascending order on condition 2; connect just the letters in alphabetical order on condition 3; alternate between connecting numbers and letters on condition 4; and draw a line on the dotted line as quickly as possible on condition 5.	Assessing visual scanning, number sequencing, letter sequencing, number-letter switching and motor speed.
Verbal Fluency Test (VFT)	This test comprises 3 conditions, the Letter Fluency, the Category Fluency and the Category Switching. The examinee is asked to say words that begin with a specific letter (F, A and S) on condition 1; say words that belong to a designated semantic category (animals and boys' names) on condition 2 and alternate between saying words from two different semantic categories (fruits and furniture) on condition 3.	Fluent productivity in the verbal domain.
Design Fluency Test (DFT)	This test comprises 3 conditions, the Filled-Dots, the Empty-Dots and the Dot Switching. The examinee is presented rows of boxes containing an array of dots and must make as many designs as possible within one minute time limit by connecting filled dots on condition 1; by connecting unfilled dots on condition 2; and by alternating connections between filled and unfilled dots on condition 3.	Fluent productivity in the nonverbal domain.
Colour-Word Interference Test (CWIT)	This test is a variant of the 'Stroop procedure' which comprises 4 conditions, the Colour Naming, the Word Reading, the Inhibition and the Inhibition/ Switching. The examinee has to name colour patches on condition 1; read colour-words printed in black ink on condition 2; name the ink colour in which colour words are printed on condition 3; and switch back and forth between naming the dissonant ink colours and reading the conflicting words on condition 4.	Inhibition of an overlearned response and mental flexibility.
Sorting Test (ST)	This test has 2 conditions, the free sorting and the sort recognition. In free sorting, the examinee is asked to sort 6 cards into 2 groups, according to as many rules as possible. In sort recognition, the examinee has to identify and describe the correct rules.	Problem-solving, verbal and nonverbal concept formation, and flexibility of thinking on a conceptual task.
Twenty Questions Test (TQT)	The examinee is presented with a stimulus page containing 30 common objects and has to identify the target by asking the fewest number of yes/no questions.	Categorical processing and ability to use feedback to guide problem solving
Word Context Test (WCT)	The examinee has to discover the meaning of made-up words based on cues given in sentences.	Verbal abstract thinking, deductive reasoning and hypothesis testing
Tower Test (TT)	The examinee's task is to move 5 disks across 3 pegs to build a target tower in the fewest number of moves possible.	Planning, rule learning, and inhibition
Proverb Test (PT)	Unlike the other 8 D-KEFS subtests, this test is for adolescents and adults aged 16 to 89. It consists of 8 sayings that are presented into 2 conditions, free inquiry and multiple choice. The examinee has to interpret proverbs orally and select the best interpretation from 4 alternatives.	Metaphorical thinking, generating and comprehending abstract thought

Major features of the D-KEFS

For most of the D-KEFS subtests, scaled scores are converted from raw scores and have a mean of 10 and a standard deviation of 3. The D-KEFS as a whole can generate a total of 125 scores, of which 42 are "primary" performance measures and 83 are "optional" measures. The optional scores are additional measures which include error, contrast, accuracy and time-interval measurement (Delis, Kaplan & Kramer, 2001a). These kinds of information are essential to provide a more comprehensive assessment of executive functioning. Particular strategies and error types committed, for instance, provide qualitative information which is important in the evaluation of an individual's neuropsychological functioning profile. In addition to this, the use of contrast measures also facilitates the "process" interpretation of the scores which is useful in terms of identifying any neurocognitive mechanisms underlying poor performance under different subtest conditions (Homack, Lee & Riccio, 2005; Swanson, 2005). This process-orientated interpretation is one of the features that discriminates the D-KEFS from most other executive function tests. It aims to isolate the relative contributions of more fundamental cognitive skills (e.g., language, visuo-perception) from the higher-level cognitive functions (e.g., cognitive flexibility, problem solving). The examiner can thus determine and assess how these component processes might have influenced the higher-level executive performance when a person performs poorly on a task (Swanson, 2005). This essentially helps clinicians to tease out executive dysfunctions from fundamental cognitive deficits. As a result, the D-KEFS offers the promise of a more effective evaluation of executive functioning than many other tests of executive functioning that have failed to differentiate non-executive processing.

The inclusion of switching conditions is also another major characteristic of the D-KEFS. These new switching conditions have been added to several of the D-KEFS subtests, including the Trail Making Test, Colour-Word Interference Test, the Verbal Fluency Test and the Design

Fluency Test. These switching procedures require the participants to switch between two cognitive set of conditions or categories. In the Verbal Fluency switching task, for instance, the participant is asked to shift back and forth between naming as many fruits and as many pieces of furniture as he can. These switching tasks can thus increase the executive processing demands of the tests, thereby maximising the sensitivity to detect subtle executive function deficits (Swanson, 2005).

What do we know about the validity of the D-KEFS?

Whilst the D-KEFS technical manual (Delis, Kaplan, & Kramer, 2001b) reported that the D-KEFS is a promising test in the measurement of executive functioning in terms of its psychometric properties, Schmidt (2003) has argued that the manual contains insufficient independent evidence to support the validity of the test. In response, Delis and his colleagues have rebutted this criticism by pointing out that much of the validity data for the D-KEFS has subsequently been published in peer-reviewed journals within the mainstream neuropsychology literature (Delis, Kramer, Kaplan & Holdnack, 2004). Previously, four papers have provided an in-depth narrative review of the D-KEFS, which include the detailed descriptions of the test, its administration, scoring and interpretation, standardisation, as well as the technical characteristics of various D-KEFS subtests (Baron, 2004; Homack, Lee & Riccio, 2005; Swanson, 2005; Shunk, Davis & Dean, 2006). Despite the fact that the D-KEFS was reported as a promising clinical instrument for the assessment of executive functioning in these reviews, the psychometric properties of the various subtests were not fully documented. Chesters (2008) further reviewed the validity of the D-KEFS by synthesising specific evidence in a variety of clinical populations. Although this reviewer concluded that D-KEFS is a useful tool in the measurement of executive functions in a range of both clinical and educational settings, he pointed out that the Proverb Test appeared to have received less empirical attention than other

more commonly administered D-KEFS subtests such as the Trail Making, Verbal Fluency and Colour-Word Interference tests. More work might therefore need to be done in future study to establish the utility of the Proverb Test.

Objectives of this systematic review

Although these review papers have demonstrated a substantial body of evidence in support of the effectiveness of the D-KEFS with regard to its ability to identify executive dysfunctions among different clinical groups, they have not been updated for at least a decade and the results do not reflect contemporary evidence. To the best of the author's knowledge, there has been no systematic review to date that has been conducted to evaluate the evidence base relating to the clinical usefulness of the D-KEFS. Given the value of assessing the executive functions in people with acquired neurological conditions, the objectives of the present review is to update and extend the existing neuropsychological literature on the D-KEFS, specifically aiming to evaluate the evidence about its clinical usefulness with respect to the identification of the impairment in executive functioning between people suffering from acquired neurological pathologies and healthy controls or between different clinical groups.

Research questions

In order for a D-KEFS subtest to be considered suitable for use in a neuropsychological context it would seem important that there should be sufficient evidence to address the following questions.

1. Would participants with acquired neurological conditions show significant executive impairment (as reflected from their mean scores or the number and type of errors), to be discernible from demographically matched healthy controls?

2. To what extent does the D-KEFS show sensitivity and specificity to known patterns of neural deficits in different neurological pathologies?

METHODOLOGY

In order to gain a more accurate indication of the clinical usefulness of the D-KEFS as a clinical assessment tool for identifying executive dysfunctions in various acquired neurological conditions, a systematic review was conducted. A comprehensive search for relevant studies using explicit and transparent search terms was therefore employed in this review. Any studies that assessed the diagnostic accuracy and discriminative ability of the D-KEFS to distinguish neurologically impaired individuals from healthy participants, or differentiate between various groups of neurological conditions were synthesised and critically analysed. The methodology and reporting of this review followed the 'PRISMA' (Preferred Reporting Items for Systematic reviews and Meta-Analysis) statement (Moher, Liberati, Tetzlaff & Altman, 2009).

Eligibility criteria

Articles were included if they met the following criteria for eligibility:

1. Studies that utilised the D-KEFS subtest(s) as a primary measure of executive functioning

2. Studies that reported the clinical usefulness of the D-KEFS subtest(s) with regard to its diagnostic accuracy or ability to identify executive function impairment in people with various acquired neurological pathologies

3. Studies that included clinical populations with diagnoses of acquired neurological conditions

4. Studies published in peer-reviewed journals and in the English language

5. Participants aged between 8 and 89 due to the D-KEFS being normed for use with 8-89 years old individuals (with the exception of the Proverb Test)

Publication within a peer reviewed journal was included to ensure high quality and validity of research papers were being reviewed. Review articles, studies that involved just one patient, research papers whose aims did not encompass the evaluation of the clinical usefulness of the D-KEFS, and/ or reporting data regarding clinical groups other than acquired neurological conditions (e.g., neurodevelopmental disorders, psychiatric illnesses or intellectual disability) were all excluded in this review.

Search strategy

The literature search covered from 2001 (i.e. the publication date of the D-KEFS) through to December 2018. To identify all eligible studies, three computerised databases of PsycINFO, MEDLINE(R) and EMBASE were used in this review. Combinations of search terms were used to carry out the search: "dysexecutive OR executive dysfunction OR executive function* OR frontal lobe function*" AND "Delis Kaplan Executive Function* System OR DKEFS" AND "sensitivity OR specificity OR clinical utility OR clinical usefulness OR diagnostic OR validation OR validity" AND "acquired neurological disorders". Specific search terms are listed in Table 2.

Table 2: Databases and search terms

Database	Search terms	Number of articles identified
PsycINFO	"dysexecutive OR executive dysfunction OR executive function* OR frontal lobe function*" AND "Delis Kaplan Executive Function* System OR DKEFS" AND "sensitivity OR specificity OR clinical utility OR clinical usefulness OR diagnostic OR validation OR validity" AND "acquired neurological disorders" (explode: nervous system disorders/ neurology/ measurement/ traumatic brain injury/ cognitive impairment/ epilepsy)	301
MEDLINE(R)	"dysexecutive OR executive dysfunction OR executive function* OR frontal lobe function*" AND "Delis Kaplan Executive Function* System OR DKEFS" AND "sensitivity OR specificity OR clinical utility OR clinical usefulness OR diagnostic OR validation OR validity" AND "acquired neurological disorders" (explode: human/ magnetic resonance imaging/ multiple sclerosis/ epilepsy/ temporal lobe)	51
EMBASE	"dysexecutive OR executive dysfunction OR executive function* OR frontal lobe function*" AND "Delis Kaplan Executive Function* System OR DKEFS" AND "sensitivity OR specificity OR clinical utility OR clinical usefulness OR diagnostic OR validation OR validity" AND "acquired neurological disorders" (explode: neurologic disease/ multiple sclerosis/ traumatic brain injury/ human)	104

Study selection

Studies eligible for inclusion were evaluated in three phases. In the first phase, all study titles were screened for relevance according to the pre-specified selection criteria. Titles which had been excluded by the author were not considered further. Following this, abstracts of all the remaining papers were then reviewed. Full text articles were only obtained and examined if the abstracts did not provide sufficient information about the inclusion criteria. Finally, full texts of the remaining articles were retrieved and assessed for eligibility. The reference lists of all included articles were also reviewed to identify any possible additional articles that met the eligibility criteria.

After applying the search limits, the three databases altogether yielded 456 publications. All titles were then screened according to the selection criteria, 341 of which were subsequently excluded. Next, abstracts of the remaining 54 articles were reviewed for eligibility. This

inspection led to the exclusion of a further 29 articles. Reasons for exclusion included the following: the D-KEFS was not principally used to assess the executive functioning; single-case study; papers which did not encompass the evaluation of the clinical usefulness of the D-KEFS in executive performance; studies in which clinical diagnoses other than acquired neurological conditions were reported; and articles in which only a normal population was recruited without any group comparisons in terms of the D-KEFS performances. As a result, 25 studies fulfilling the eligibility criteria were selected. During the inspection, the reference lists of the included articles were also hand-searched for possible relevant studies. This resulted in the inclusion of 7 additional studies. In total, thirty-two studies were identified in this systematic review.

Identification

Records identified through PsycINFO, MEDLINE(R) and EMBASE databases searching (n = 456)

Number of duplicates removed (n = 61)

Screening

Records screened for relevance (n = 395)

Records excluded by the title and/ or abstract (n = 341)

Eligibility

Full-text articles assessed for eligibility (n = 54)

Full-text articles excluded, with reasons, (n = 29):

- 5 articles in which executive functions were not principally measured by the D-KEFS
- 3 review papers
- 4 were not peer-reviewed journals
- 1 single case study
- 5 articles that did not evaluate the clinical usefulness of the D-KEFS for its discriminative ability
- 3 studies included other diagnoses such as neurodevelopmental disorders (ASD, ADHD, cerebral palsy, learning disability), psychiatric illnesses (depression, schizophrenia) and other diagnoses (substance abuse)
- 8 articles in which the study participants were normal population

Included

Additional studies identified from reviewing reference lists of included papers (n = 7)

Studies included from reading through texts (n = 25)

Studies included in qualitative narrative synthesis (n = 32)

Figure 1: Flow chart depicting the review process and application of the eligibility criteria

Data extraction and analysis

Specific information was extracted from each eligible study by the author. Fields of interest included: year of publication; country where the study was conducted; source of recruitment; study aims; number of participants; age and demographic characteristics of participants; neurological condition diagnosed; D-KEFS subtest(s) that was/ were used; study methodology; assessment of outcomes; key findings and limitations of the respective study.

In order to answer the research questions outlined above, the following outcome measures were extracted and synthesised for analysis:

1. Group comparisons of the D-KEFS performance in executive functioning

Studies comparing mean D-KEFS performance scores from participants with acquired neurological conditions and from healthy control participants were examined. Findings reporting the number and type of errors committed were also extracted as part of the analysis. It was considered that all of these could provide essential evidence supporting the validity of the D-KEFS in the form of identifying impairment in individuals with neurological pathologies.

2. Diagnostic accuracy metrics to quantify the discriminative property of the D-KEFS

Diagnostic accuracy provides evidence on how well an instrument can correctly identify or rule out a diagnosis, i.e. differentiating the diseased from those who are healthy (Wong & Lim, 2011). Such a discriminative property can be assessed and quantified by the measures of diagnostic accuracy such as likelihood ratios, sensitivity, specificity, the area under the Receiver Operating Characteristic (ROC) curve and classification accuracy. Some measures are used to assess the discriminative ability of a given test, whereas others are more related to the estimation of its predictive power (Irwig,

Bossuyt, Glasziou, Gatsonis & Lijmer, 2002). Therefore, studies reporting the diagnostic accuracy metrics of the D-KEFS are considered useful to illustrate its ability to detect or exclude executive dysfunctions, or to differentiate individuals with various neurological pathologies from normal controls.

3. Correlation of D-KEFS performance with brain regions or with specific neural patterns in various acquired neurological conditions

Studies addressing the relationship between brain areas or neuroanatomical substrates and the D-KEFS performances were also identified. By exploring such relationship, the ability of the D-KEFS in differentiating neurologically impaired patients associated with damage in a specific brain region can generally be evaluated.

Risk of bias in individual studies

The risk of bias in each empirical study was assessed according to a set of quality appraisal criteria1, which are fully described in Appendix I. Within each of the ten domains of the quality appraisal criteria, a series of guiding questions were set out to elicit information about the methodological features reported in each study which were relevant to risk of bias. Biases were judged as "Low risk bias", "High risk bias" or "Unclear risk bias" accordingly on the basis of the information reported. The judgement of "Low risk bias" (Green rating) was made when there was sufficient information to suggest that a plausible bias was unlikely to seriously alter the study results. A full score of two points were awarded for a low risk of bias.

---

1 The criteria for assessing the quality of the included studies in this review was based on a number of papers. The STROBE statement (Strengthening the Reporting of Observational Studies in Epidemiology) proposed by von Elm et al (2007) takes a particular view on what considers as a good reporting of observational studies. This 22-item checklist was selected because it was established to provide guidance for reviewers to critically appraise published research articles (von Elm et al, 2007). Additionally, the developers of the STROBE statement recommended that this checklist is best used in conjunction with an accompanying paper (Vandenbroucke et al, 2007), in which the meaning and rationale for each checklist item were explained and elaborated. Apart from this, another paper "Step-by-Step Guide to Critiquing Quantitative Research" (Coughian, Cronin & Ryan, 2007) was also used in determining quality criteria upon which this systematic review is based. Based on the ideas set out from these papers, the author established a set of detailed quality appraisal criteria.

However, when there was suggestive evidence that a plausible bias was likely to have the capacity to influence the study results and weaken the confidence in the findings. A "High risk bias" (Red rating) was judged and no score was given in this instance. For any bias which was unclear as to whether it would affect the study outcome but it was considered that the reader should be aware of the bias when interpreting the results, or there was lack of information to indicate that an important risk of bias might exist, this was categorised as an "Unclear risk bias" (Amber rating). One point was awarded for this rating. At the end of this process each paper should be assigned a score out of 20 to indicate its quality. Table 3 below provides a summary of the methodological quality of the studies reviewed.

Table 3 – Review of methodological quality of the included studies

Studies reviewed	Background/ Rationale & Study Objectives	Study Design	Study Setting & Participants	Sample Size	Attrition	Data Collection	Analysis/ Results	Potential Bias	Limitations	Interpretation/ Generalisability	Quality Score (out of 20)
Anderson et al (2017)	GREEN	GREEN	GREEN	GREEN	AMBER	GREEN	GREEN	AMBER	GREEN	AMBER	16
Baldo et al (2004)	AMBER	GREEN	GREEN	RED	RED	GREEN	GREEN	AMBER	GREEN	AMBER	12
Baldo et al (2001)	GREEN	GREEN	RED	RED	RED	GREEN	GREEN	RED	GREEN	AMBER	10
Carey et al (2008)	GREEN	GREEN	GREEN	GREEN	RED	AMBER	GREEN	GREEN	GREEN	GREEN	17
Clark et al (2012)	AMBER	GREEN	AMBER	GREEN	RED	AMBER	GREEN	GREEN	GREEN	GREEN	14
Faber et al (2016)	GREEN	GREEN	RED	AMBER	RED	GREEN	GREEN	GREEN	GREEN	GREEN	12
Fine et al (2009)	GREEN	GREEN	RED	RED	RED	GREEN	AMBER	RED	GREEN	GREEN	10
Fine et al (2008)	GREEN	GREEN	GREEN	GREEN	RED	GREEN	GREEN	GREEN	GREEN	GREEN	12
Gansler et al (2017)	GREEN	AMBER	GREEN	GREEN	RED	GREEN	GREEN	AMBER	AMBER	GREEN	14
Ghawami et al (2017)	GREEN	GREEN	GREEN	RED	RED	GREEN	GREEN	RED	GREEN	RED	9
Heled et al (2012)	GREEN	GREEN	AMBER	GREEN	AMBER	GREEN	GREEN	AMBER	GREEN	AMBER	15
Houston et al (2005)	GREEN	GREEN	GREEN	AMBER	RED	GREEN	GREEN	GREEN	GREEN	GREEN	17
Huey et al (2009)	GREEN	GREEN	AMBER	AMBER	AMBER	GREEN	GREEN	AMBER	AMBER	GREEN	15
Kaiser et al (2013)	GREEN	GREEN	GREEN	RED	RED	GREEN	GREEN	AMBER	AMBER	GREEN	12
Keifer et al (2013)	GREEN	GREEN	AMBER	RED	RED	GREEN	GREEN	GREEN	GREEN	AMBER	14
Keil et al (2005)	GREEN	GREEN	RED	RED	RED	GREEN	GREEN	RED	AMBER	AMBER	10
Kramer et al (2007)	GREEN	AMBER	GREEN	GREEN	RED	GREEN	GREEN	GREEN	GREEN	GREEN	14
Luton et al (2010)	GREEN	AMBER	AMBER	AMBER	RED	AMBER	GREEN	AMBER	GREEN	GREEN	13
McDonald et al (2008)	GREEN	GREEN	GREEN	GREEN	RED	GREEN	GREEN	GREEN	GREEN	GREEN	15
McDonald et al (2005a)	GREEN	GREEN	GREEN	GREEN	RED	GREEN	GREEN	AMBER	GREEN	GREEN	14
McDonald et al (2005b)	GREEN	GREEN	GREEN	GREEN	RED	GREEN	GREEN	GREEN	GREEN	GREEN	15
McDonald et al (2005c)	GREEN	GREEN	GREEN	GREEN	RED	GREEN	GREEN	GREEN	GREEN	GREEN	15
Meoded et al (2013)	AMBER	GREEN	RED	GREEN	RED	GREEN	GREEN	GREEN	GREEN	AMBER	10
Nutter-Upham et al (2008)	GREEN	GREEN	GREEN	GREEN	AMBER	GREEN	GREEN	GREEN	GREEN	GREEN	19
Parmenter et al (2007)	GREEN	GREEN	AMBER	GREEN	RED	GREEN	GREEN	AMBER	GREEN	GREEN	15
Parrish et al (2007)	AMBER	GREEN	GREEN	GREEN	RED	GREEN	GREEN	GREEN	GREEN	GREEN	16
Pulsipher et al (2009)	GREEN	GREEN	GREEN	GREEN	AMBER	GREEN	GREEN	GREEN	GREEN	GREEN	18
Ryes et al (2017)	GREEN	GREEN	AMBER	GREEN	RED	GREEN	GREEN	AMBER	AMBER	GREEN	15
Strong et al (2011)	GREEN	GREEN	AMBER	GREEN	RED	AMBER	GREEN	AMBER	GREEN	GREEN	14
Wetter et al (2005)	GREEN	GREEN	GREEN	GREEN	RED	GREEN	GREEN	AMBER	GREEN	GREEN	14
Yochim et al (2009)	GREEN	GREEN	AMBER	RED	RED	GREEN	GREEN	AMBER	GREEN	AMBER	12
Yochim et al (2007)	AMBER	GREEN	RED	RED	RED	GREEN	GREEN	RED	AMBER	AMBER	9

GREEN – Low risk of bias. AMBER – Unclear risk of bias. RED – High risk of bias.

In total, thirty-two papers were scored for methodological quality, with ratings ranging from 19 (Nutter-Upham et al., 2008) to 9 (Ghawami et al., 2017; Yochim et al., 2007). Instead of discussing the quality of each paper in turn, the critical review in this section is focused specifically on areas with respect to issues of sampling, data collection, potential bias and interpretation of the findings.

With regard to sampling, none of the studies reported power calculation to estimate the necessary sample size. Of particular note is the number of studies which used relatively small samples. Six studies had fewer than 15 participants in both the clinical and control groups (Kaiser et al., 2013; Yochim et al., 2009; Yochim et al., 2007; Keil et al., 2005; Baldo et al., 2004; Baldo et al., 2001), and one study only included nine participants in the clinical group (Ghawami et al., 2017). According to the criteria proposed by von Elm et al. (2007), a group size of 15 or less is considered weak in comparison studies. Given that the power of a study depends on its sample size, studies with small sample size are less likely to have sufficient power to detect a true effect and this in turn reduces the ability to draw reliable conclusions from the study findings. Eleven of the reviewed studies were rated 'Green' in this regard, indicating that they had large enough samples to ensure a significant power. Conversely, 'Red' ratings were given for those which had insufficient sample size.

Furthermore, nearly all studies failed to give detailed information regarding the participation rate of the study subjects or report the numbers of individuals at each stage of the study. Only one study did so by indicating the numbers of participants who were initially approached, how many had refused to participate and how many were actually included in the study (Heled et al., 2012). These types of data are useful to determine the extent to which the representativeness of the

sample has been affected and whether the comparison groups are still comparable in characteristics.

Despite the fact that most of the reviewed studies clearly specified their eligibility criteria for the identification of the target population, six studies included no information about where the clinical and control groups were drawn from and how they were recruited (Faber et al., 2016; Yochim et al., 2007; Keil et al., 2005; Baldo et al., 2001; Meoded et al., 2013; Fine et al., 2009). Thus, it is difficult to determine how well the actual study participants matched the target population defined in the study question or the extent to which the comparison groups were comparable with respect to certain characteristics other than the disease of interest. This can potentially lead to sources of selection bias and confounding when the groups are not comparable. Additionally, in one study all participants recruited for the comparison groups were male (Ghawami et al., 2017). The inclusion of male only samples probably threatened the ability to make generalisation from the study results to the TBI population. Therefore, the findings of this study need to be interpreted in caution. In addition, two studies chose first-degree cousins as their controls (Parrish et al., 2007; Pulsipher et al., 2009) in which the rationale for their selection had been clearly stated. Finally, another important consideration in relation to data collection is the fact that only five studies clearly stated that the D-KEFS was administered and recorded by highly trained technicians, postdoctoral fellows or research assistants in accordance with standard manual instructions (Anderson et al., 2017; Yochim et al., 2009; Parmenter et al., 2007; Keifer & Tranel, 2013; Nutter-Upham et al., 2013). Only one study reported blinding of examiners during administration and scoring (Keifer & Tranel, 2013).

RESULTS

Study characteristics

The majority of the studies were conducted in the United States of America (n = 30); one in the

Iran (n = 1) and one in the Israel (n = 1). All studies evaluated the executive functioning of

individuals with various acquired neurological conditions using D-KEFS. The mean ages for

clinical groups ranged between 10.1 and 75 years, whereas those for healthy controls were

reported as between 12.7 and 77.5 years. This indicates that the study samples were composed

of children, adults and elderly. For most of the studies, clinical and control groups were matched

for age, gender and education level. However, there were two studies which did not mention any

matching criteria or describe how the clinical and control groups were matched (Fine et al.,

2009; Kramer et al., 2007). One study used all male participants (Ghawami et al., 2017),

whereas the remaining studies included 33% − 79% male in clinical groups and 28% − 81%

male in control groups. Furthermore, regarding study design, thirty of the studies were cross-

sectional in nature, which employed case/ control comparisons at a specific point in time for

their methodology, whereas two were prospective longitudinal studies to evaluate the predictive

validity of the D-KEFS. The sample sizes in the clinical group for those cross-sectional studies

ranged from a relatively small sample size of nine (Ghawami et al., 2017) to the largest of one

hundred and twenty-four (Gansler et al., 2017). The control group size ranged from nine (Fine et

al., 2009) to sixty-five (Strong et al., 2011). Five studies did not include any healthy controls, of

which three aimed to compare group differences in D-KEFS performance across different

neurological conditions (Keifer & Tranel, 2013; Gansler et al., 2017; Kaiser et al, 2013) and two

were longitudinal studies investigating the utility of D-KEFS to predict cognitive decline in

normal functioning older adults (Clark et al., 2012; Fine et al., 2008).

In terms of settings and locations from which the samples were drawn, most of the studies recruited participants in a range of settings. Twenty-six studies recruited their clinical samples from a diversity of specialist neurology services consisting of day, out-patient, in-patient settings, rehabilitative facilities, university hospitals or clinics. In contrast, six did not clearly state the sources of recruitment. Meanwhile, healthy participants were locally recruited from the community for the majority of the studies. Notably, two of them reported using family members of the study participants as healthy controls; six obtained the control groups from the D-KEFS normative database; ten did not report how and where the study controls were recruited; and five did not include a healthy comparison group in their study designs. With regard to the nature of clinical groups, a broad range of neurological conditions were studied. A total of five studies examined the executive functioning of individuals with Traumatic Brain Injury (TBI); six focused on patients with focal brain lesions; thirteen conducted analysis on different types of neurodegenerative disorders; and eight focused on epilepsy. The selection of the D-KEFS subtests varied across all reviewed studies. One study used the entire nine stand-alone D-KEFS subtests for the measurement; fifteen studies reported administering of several D-KEFS subtests as their assessment tools; and sixteen studies exclusively focused on a single particular D-KEFS subtest, of which three focused on the Colour-Word Interference Test, one on the Word Context Test, two on the Design Fluency Test, two on the Trail Making Test, three on the Sorting Test, two on the Tower Test, one on the Twenty Questions Test, one on the Proverb Test and one on the Verbal Fluency Test. The summary information of the demographic characteristics of the study participants is presented in Table 4.

Table 4 – Demographic characteristics of participant population in the reviewed studies under different categories of acquired neurological condition

Study	Country	Sample (size & type)				Mean age (years): Clinical group(s)/ Healthy controls	Healthy controls matched by	Neuro-imaging data included
		Clinical group(s)	Gender M (%); F (%)	Healthy controls	Gender M (%); F (%)			
Traumatic Brain Injury								
Anderson et al, 2017	US	Sample 1: 128 with TBI	73(57%); 55(43%)	-	-	38.1/ -	-	✗
		Sample 2: 28 with moderate-to-severe TBI	22(79%); 6(21%)	56	44(79%); 12(21%)	35.5; 35.7/ 35.6	Age, Gender, Ethnicity, Education	
		28 with mild-uncomplicated TBI (Selected from sample 1)	22(79%); 6(21%)					
Ghawami et al, 2017	Iran	9 with frontal lesions	All male	45	All male	21.4/ 23.9	Age, Gender, Race, Religion, Education, Socio-economic status	✗
Faber et al, 2016	US	21 with moderate-to-severe TBI sustained between 5-14 years ago	15(71%); 6(29%)	16	13(81%); 3(19%)	14.05/ 13.9	Age, Gender, Ethnicity, Socio-economic status, Handedness	✓
Heled et al, 2012	Israel	29 with severe TBI	20(69%); 9(31%)	38	13(34%); 25(66%)	25.5/ 25.2	Age, Education	✗
Strong et al, 2011	US	65 with complicated mild-severe TBI	45(69%); 20(31%)	65	Not mentioned	32/ not mentioned	Age, Gender, Ethnicity, Education	✗
Total participants:		Number of participants in clinical groups: 252		Number of healthy controls: 220				
Frontal Brain Lesions								
Keifer et al, 2013	US	45 with frontal lesions (13 vmPFC; 14 dlPFC; 18 non-frontal)	26(58%); 19(42%)	-	-	60; 58.2; 60.8/ -	-	✓
Yochim et al, 2009	US	12 with lateral PFC lesions	9(75%); (25%)	12	8(67%); 4(33%)	63.2/ 65.8	Age, Education, Socio-economic status	✗
Yochim et al, 2007	US	12 with lateral PFC lesions	8(67%); 4(33%)	11	7(64%); 4(36%)	65.5/ 68.1	Age, Education	✗
Keil et al, 2005	US	12 with prefrontal cortex lesions	8(67%); 4(33%)	12	8(67%); 4(33%)	65.1/ 63.7	Age, Education	✗
Baldo et al, 2004	US	12 with prefrontal cortex lesions	8(67%); 4(33%)	12	8(67%); 4(33%)	63.1/ 64.1	Age, Education	✗
Baldo et al, 2001	US	11 with focal frontal lesions	7(64%); 4(36%)	11	8(73%); 3(27%)	65.8/ 68.1	Age, Education	✗
Total participants:		Number of participants in clinical groups: 104		Number of healthy controls: 58				

Study	Country	Sample (size & type)				Mean age (years): Clinical group(s)/ Healthy controls	Healthy controls matched by	Neuro-imaging data included
		Clinical group(s)	Gender M (%); F (%)	Healthy controls	Gender M (%); F (%)			
				Neurodegenerative Conditions				
Gansler et al, 2017	US	124 with bvFTD; 34 with PPA; 85 with CBS	66(53%); 58(47%) 15(44%); 19(56%) 45(53%); 40(47%)	-	-	62; 62.1; 65.4/ -	-	✓
Kaiser et al, 2013	US	11 with bvFTD; 10 with AD	7(64%); 4(36%) 5(50%); 5(50%)	-	-	60.55; 57/ -	Age, Gender, Education, Ethnicity, MMSE	✓
Meoded et al, 2013	US	25 with PLS; 25 with ALS	13(52%); 12(48%) 12(48%); 13(52%)	17	9(53%); 8(47%)	56.3; 57.7/ 59.2	Age	✓
Clark et al, 2012	US	71 initially non-demented older adults (51 cognitively normal; 20 MCI)	Not mentioned	-	-	77.3/ -	-	✗
Fine et al, 2009	US	19 with AD; 25 with FTD; 13 with SD; 12 with PNFA; 11 with PSP or possible PSP	Not mentioned	9	Not mentioned	57.4; 58.6; 63.1; 63.7; 65.7/ 57	Did not mention	✓
Huey et al, 2009	US	51 with bvFTD; 50 with CBS	51(51%); 50(49%)	14	7(50%); 7(50%)	59.6; 65.6/ 60.1	Age	✓
Carey et al, 2008	US	44 with FTD; 30 with AD	32(73%); 12(27%) 18(60%); 12(40%)	27	11(41%); 16(59%)	61.5; 63.3/ 60.2	Age, Education, Ethnicity	✓
Fine et al, 2008	US	24 initially non-demented older adults	11(46%); 13 (54%)	-	-	75.9/ -	-	✗
Nutter-Upham et al, 2008	US	37 with amnestic MCI; 37 with cognitive complaints	17(46%); 20(54%) 16(43%); 21(57%)	33	11(33%); 22(67%)	71.8; 73.4/ 71.1	Age, Education, Gender	✗
Kramer et al, 2007	US	16 with AD; 30 with FTD; 19 with SD	Not mentioned	36	Not mentioned	60.8; 58; 61.9/ 64.4	Did not mention	✓
Parmenter et al, 2007	US	111 with MS	33(30%); 78(70%)	46	13(28%); 33(72%)	44.8; 43.8	Age, Gender, Race	✓
Houston et al, 2005	US	24 with APOE-e4 (genetic risk factor of AD)	13 (54%); 11(56%)	28 without APOE-e4	8(29%); 20(71%)	75/ 77.3	Age, Education, Gender, Global cognitive status	✗

27

Study	Country	Sample (size & type)				Mean age (years): Clinical group(s)/ Healthy controls	Healthy controls matched by	Neuro-imaging data included
		Clinical group(s)	Gender M (%); F (%)	Healthy controls	Gender M (%); F (%)			
Neurodegenerative Conditions								
Wetter et al, 2005	US	22 with APOE-e4 (genetic risk factor of AD)	11(50%); 11(50%)	29 without APOE-e4	9(31%); 20(69%)	75/ 77.5	Age, Education, Gender, Global cognitive status	✗
Total participants:		Number of participants in clinical groups: 865 (excluding those in Clark et al, 2012 & Fine et al, 2008)		Number of healthy controls: 239				
Epilepsy								
Reyes et al, 2017	US	32 with TLE;	17(53%); 15(47%)	24	11(46%); 13(54%)	38.2/ 36.5	Age, Gender	✓
Luton, Burns & DeFilippis, 2010	US	20 with FLE	13(65%); 7(35%)	20	13(65%); 7(35%)	12.4/ 12.7	Age, Gender	✗
Pulsipher et al, 2009	US	20 with recent onset juvenile myoclonic epilepsy; 12 with recent onset BCECTS	17(53%); 15(47%)	51	22(43%); 29(57%)	15.5; 10.1 /13.2	Gender, Intelligent quotient	✓
McDonald et al, 2008	US	22 with FLE; 20 with TLE	9(41%); 13(59%) 7(35%); 13(65%)	23	10(43%); 13(57%)	38.9; 37.9/ 36.9	Age, Education, Gender	✗
Parrish et al, 2007	US	53 with new onset epilepsy	31(58%); 22(42%)	50	23(46%); 27(54%)	12.7/ 12.7	Age, Sex, Education	✗
McDonald et al, 2005a	US	22 with FLE; 20 with TLE	16(38%); 26(62%)	23	10(43%); 13(57%)	36.8; 37.9/ 36.9	Age, Education, Gender	✗
McDonald et al, 2005b	US	23 with FLE; 20 with TLE	Not mentioned	23	Not mentioned	36.8; 37.9/ 36.9	Age, Education, Gender	✗
McDonald et al, 2005c	US	23 with FLE; 20 with TLE	17(40%); 26(60%)	23	10(43%); 13(57%)	36.8; 37.9/ 36.9	Age, Education, Gender	✗
Total participants:		Number of participants in clinical groups: 307		Number of healthy controls: 237				

Note: TBI = Traumatic Brain Injury; bvFTD = behavioural variant Frontotemporal Dementia; PPA = Primary Progressive Aphasia; CBS = Corticobasal Syndrome; vmPFC = ventromedial Prefrontal Cortex; PLS = Primary Lateral Sclerosis; ALS = Amyotrophic Lateral Sclerosis; dlPFC = dorsolateral Prefrontal Cortex; PFC = Prefrontal Cortex; OFC = Orbito-Frontal Cortex; FTD = Frontotemporal Dementia; AD = Alzheimer's Dementia; MS = Multiple Sclerosis; MCI = Mild Cognitive Impairment; SD = Semantic; Dementia; PNFA = Progressive Non-Fluent Aphasia; PSP = Progressive Supranuclear Palsy; FLE = Frontal Lobe Epilepsy; TLE = Temporal Lobe Epilepsy; BCECTS = Benign Childhood Epilepsy with Centro-Temporal Spikes; APOE-e4 = e4 allele of Apolipoprotein; MMSE = Mini-Mental State Examination

The clinical usefulness of the D-KEFS

The thirty-two studies included in this review collectively examined the clinical usefulness of the D-KEFS. Due to the heterogeneity of the studies, the following section was organised according to types of neurological pathology. An overview of the reviewed studies summarising the study aims, what D-KEFS measures were used, main findings and the limitations is illustrated in Appendix II.

Traumatic Brain Injury, TBI (n = 5)

All five studies reported that participants with TBI performed statistically worse than healthy controls on different D-KEFS measures. Anderson et al. (2017) found that people with moderate-to-severe TBI performed significantly worse on both the D-KEFS CWIT Inhibition/ Switching and VFT Category Switching tasks than both the mild-uncomplicated TBI and control groups. The combined subtests performed well in identifying TBI group with cognitive impairment but the model lacked sufficient discriminative power when it was used to distinguish people with moderate-to-severe TBI from those with mild-uncomplicated TBI and healthy participants. Ghawami et al. (2017) compared a group of people with focal frontal contusions to a demographically matched control group on 6 D-KEFS subtests: TMT, VFT, DFT, ST, TQT and TT. The TBI group showed substantial executive dysfunctions on all measures with large effect sizes. Of particular note is that these patients' performances still remained significantly impaired on switching conditions even when fundamental conditions were controlled for. In addition, the TBI patients also committed considerably more errors than the controls, qualitatively showing a tendency towards more repetition/ perseverative errors. For Heled et al. (2012), the study's results indicated that the severe TBI group performed considerably worse than healthy controls across 5 dimensions of the D-KEFS ST, of which the "attempted sorts" was

the most sensitive measure to distinguish group difference with it also being one of the best predictors of group classification. Strong et al. (2011) investigated the performance of patients with complicated mild-severe TBI on two D-KEFS subtests: VFT and DFT. Significant group differences, with very small effect sizes, were found on the Letter Fluency and Category Switching subtests but not on any of the DFT subtests. The combined Letter Fluency and Category Switching tasks also demonstrated a suboptimal classification accuracy when identifying people with TBI from controls. Finally, Faber et al. (2016) examined the executive functioning of children with early moderate-to-severe TBI using D-KEFS. This study's results indicated that the TBI group exhibited significantly decreased structural integrity of brain tissues compared to the control group and performed poorly on the VFT Letter Fluency, CWIT Inhibition and Inhibition/ Switching tasks as well as committing more errors on the inhibition task.

Focal Brain Lesions (n = 6)

Findings of all six studies suggest that the D-KEFS performance is adversely affected by frontal lobe lesions. Firstly, a study by Keifer & Tranel (2013) evaluated the executive functioning of three patient groups with focal lesions in the dorsolateral prefrontal cortex (dlPFC), ventromedial prefrontal cortex (vmPFC) and non-frontal (NF) regions using the entire D-KEFS. Patients with dlPFC brain lesions showed significant impairment in performance on six D-KEFS primary measures than patients with vmPFC and NF lesions. None of the D-KEFS measures could differentiate between the vmPFC and NF groups, however, all group differences on the D-KEFS became non-significant when the effects of intelligence and processing speed were taken into account. Results also suggest that the three groups did not particularly differ in their performance on the D-KEFS switching paradigms compared to the traditional counterparts.

Furthermore, Yochim and his colleagues have conducted two studies investigating the performance of patients with frontal lobe lesions on the D-KEFS TT and TMT. Yochim et al. (2009) showed that patients with frontal lobe lesions completed fewer towers, spent longer on each move and committed more rule violations on the D-KEFS TT than the controls. Of particular note is that all participants with frontal lesions committed two or more rule violations. This measure, therefore, demonstrated good sensitivity and specificity in identifying people with frontal lesions. Yochim et al. (2007) also found that the frontal lesion group performed slower than the control group on the D-KEFS TMT, particularly on the Letter Sequencing and the Number-Letter Switching tasks. In addition, patients with frontal lesions performed disproportionately slower on the switching condition even when the performance on all baseline conditions was controlled for. Notably, this patient group also showed a propensity to commit more errors on the switching task than controls. In another study, Keil et al. (2005) found that patients with frontal lesions performed significantly more poorly than matched controls on the D-KEFS WCT. Patients were less able to generate correct responses and had to make more guesses across the 10 trials. Qualitatively, patients were also found to be less able to integrate information from prior trials and infer correct response on the next trial. Furthermore, Baldo and her colleagues have examined the executive deficits of patients with frontal lesions using D-KEFS in a series of papers. Baldo et al. (2001) found that patients with frontal lesions performed significantly worse than healthy controls in verbal and design fluency tests. The frontal lesion group produced less designs on the DFT and generated fewer correct responses on the VFT but was not particularly disproportionately disadvantaged on switching conditions. The subsequent findings of Baldo et al. (2004) also suggest that patients with frontal lesions performed worse than controls on the TQT, requiring significantly more questions to be able to identify target items across all 4 trials. Qualitatively, patients also tended to rely on ineffective strategies,

focusing more on concrete attributes (e.g., "Does it use gasoline?") and specific items (e.g., "Is it the banana?") and these restricted them from narrowing down their searches.

Neurodegenerative conditions (n = 13)

Gansler et al. (2017) compared the executive functions and dysexecutive behaviours in three groups of patients with behavioural variant frontotemporal dementia (bvFTD), primary progressive aphasia (PPA) and corticobasal syndrome (CBS) using D-KEFS and Frontal System Behaviour Scale (FrSBe) respectively. On tests of executive functions, the bvFTD group showed significant impairment in performance relative to the PPA and CBS groups on TT and ST but not on any of the VFT measures. Nevertheless, the D-KEFS alone did not perform best in discriminating bvFTD from both CBS and PPA. Additionally, neuroimaging analyses indicated that the D-KEFS composite score was distinctively associated with caudal left dorsolateral prefrontal and lateral temporo-parietal cortices.

In another study, Huey et al. (2009) compared the D-KEFS performance between patients with bvFTD and patients with CBS and determined the brain regions responsible for these executive functions. Their findings showed that the bvFTD group performed significantly worse than the CBS group on the majority of the ST, TQT, TT and VFT. Among these measures, TQT and rule violations performed best in distinguishing the clinical diagnoses of bvFTD and CBS. Significant associations were also found between D-KEFS performance and frontal regions of the brain. For instance, sorting was related to dorsolateral prefrontal cortex and performance on TQT was correlated with left anterior frontal cortex. Notably, Carey et al. (2008) found that patients with frontotemporal dementia (FTD) committed significantly more rule violations than patients with Alzheimer's dementia (AD) and healthy controls on the D-KEFS TT. Individuals

who committed five or more rule violations were found to be more than four times more likely to have a FTD diagnosis. This measure demonstrated good specificity (80%) but low sensitivity (50%). Significant correlations were also found between the rule violation and bilateral frontal brain volumes. The findings of Kaiser et al. (2013) also suggested that patients with bvFTD, in comparison with those with AD, performed significantly worse on the D-KEFS PT in both accuracy and interpretation. Specifically, the bvFTD patients' responses tended to involve more concrete interpretations whereas AD patients were more likely to respond with abstract interpretations for common proverbs. However, there were no significant group difference on uncommon proverb interpretations. Poor performance in proverb interpretation was also found correlated with anterior temporal lobe region.

Parmenter et al. (2007) examined the performance of patients with multiple sclerosis (MS) and matched controls on the D-KEFS ST and the Wisconsin Card Sorting Test (WCST). Their results showed that the MS group performed significantly worse than the control group on most dimensions of both tests. Nevertheless, only the D-KEFS "Correct sorts", "Sort description" and "Sorting repetition" could significantly discriminate between the groups when the effect of depressive symptomatology was controlled for. Moreover, neuroimaging also indicated that cognitive performances on both tests were modestly correlated with the lesion volume and extent of brain atrophy. Of further interest, Meoded et al. (2013) compared the D-KEFS performance between patients with amyotrophic lateral sclerosis (ALS) and patients with primary lateral sclerosis (PLS). Their results indicated that the two groups showed no significant difference in performance on any of the D-KEFS subtests. The D-KEFS performance in this instance was linked to diffusion properties of white matter tracts. Further analysis revealed that the factors "Fluency" and "Sorting" were associated with axial diffusivity of several white matter tracts,

suggesting that the performance of these D-KEFS tasks required the integrity of tracts connecting the frontal lobes with different regions of the brain.

Fine et al. (2009) also examined the relationships between the lobar volumes of different brain regions and performance on D-KEFS ST in a mixed sample of patients with neurodegenerative diseases compared to healthy participants. Their results indicated that the left frontal lobar volume still significantly predicted the free sorting description score, when the effects of potential moderators and head size were all controlled for. In a mixed samples of patients with different types of dementia, Kramer et al. (2007) reported that both the left and right frontal volumes were significantly associated with the switching condition of the D-KEFS DFT when the effects of global cognition, working memory and baseline conditions of the task were all taken into account. Regarding patients with amnestic Mild Cognitive Impairment (MCI), Nutter-Upham et al. (2008) showed that the MCI group scored significantly lower on most measures of VFT compared to the control group. A factor analysis of fluency tasks yielded two factors: "Switching" and "Production". Analysis revealed that the "Switching" factor performed comparatively better than the "Production" factor in discriminating MCI patients from controls.

Two studies have specifically demonstrated that performance on certain D-KEFS measures is useful in discriminating a group of normal functioning older adults with subsequent cognitive decline from those who remained stable in cognitive functioning over a period of a year (Clark et al., 2012; Fine et al., 2008). In the year prior to the cognitive decline, Clark et al. (2012) found that the decline group performed significantly worse than the no-decline group on the CWIT and VFT switching conditions. In contrast, both decliners and non-decliners performed comparably on the DFT switching, TT spatial planning and TMT switching tasks. Among these measures,

the CWIT Inhibition/ Switching condition was the strongest predictor with the largest effect size, which reliably distinguished decline from no-decline outcome over a 1-year period.

Similarly, Fine et al. (2008) also indicated that the decline group performed significantly worse than the stable group on the CWIT Inhibition/ Switching condition in the year before the decline. Of particular note is that the decline group demonstrated a higher mean CWIT discrepancy score (the contrast performance between the switching and fundamental conditions) than the stable group. Additionally, this discrepancy score before decline significantly predicted cognitive decline over the subsequent year. Moreover, another two studies have evaluated the executive functioning in a sample of normal functioning older adults genetically at risk for Alzheimer's disease and determined the relationship between the presence of this risk factor (APOE genotype) and D-KEFS performance (Houston et al., 2005; Wetter et al., 2005). When mean scores were compared between the AOPE-e4 and non-e4 groups, no statistical differences were observed on any of the D-KEFS fluency measures (Houston et al., 2005) and CWIT (Wetter et al., 2005). Houston et al. (2005), however, reported that participants with APOE-e4 tended to demonstrate a higher frequency of asymmetric cognitive performances on the verbal and design fluency conditions. Wetter et al. (2005) also found that the APOE-e4 group significantly committed more errors on the Inhibition/ Switching condition than the non-e4 group. Additionally, those who demonstrated exceptionally high error rates were all in the APOE-e4 group. More importantly, a significant association between the error rates in the Inhibition/ Switching condition and the global cognitive functioning was only found in the APOE-e4 group.

Epilepsy (n = 8)

All of the study findings suggest that individuals with different kinds of epilepsy demonstrated deficits in executive functioning as reflected by the impaired performance on different D-KEFS measures. Reyes et al. (2017) reported that patients with temporal lobe epilepsy (TLE) performed significantly worse than healthy controls on the D-KEFS VFT switching task but the two groups were indistinguishable on the CWIT switching condition. In patients with TLE, a significant correlation was also found between the reduced neurite density within inferior fronto-striatal tract and poorer CWIT Inhibition/ Switching performance. Moreover, Luton et al. (2010) compared the executive abilities of children with frontal lobe epilepsy (FLE) to normal controls and investigated the differences in neurocognitive performance between those with early and later seizure onset. The results of this study showed that the FLE group demonstrated significantly greater difficulty on D-KEFS VFT and TMT relative to the control group. Additionally, children with early seizure onset performed significantly worse than controls on Category Fluency, Category Switching and Number-Letter Switching. In contrast, however, children with later seizure onset performed on these tasks comparably to the controls. Interestingly, however, Parrish et al. (2007) reported that children with recent onset epilepsy performed worse than normal controls on D-KEFS ST and CWIT. Additionally, when the epilepsy group was split into 'at-risk' and 'low-risk' group, the 'at-risk' group performed significantly worse than the 'low-risk' group on D-KEFS ST, VFT and CWIT. In children with juvenile myoclonic epilepsy, Pulsipher et al. (2009) found that the epilepsy group performed significantly worse than healthy controls on D-KEFS Inhibition, with moderate effect size. Regression analysis also indicated that frontal and thalamic volumes were the best significant predictors of performance on all D-KEFS measures.

In patients with frontal lobe epilepsy (FLE), impaired D-KEFS performance was reported in a series of studies conducted by McDonald and her colleagues. Firstly, McDonald et al. (2005a) found that patients with FLE performed significantly slower than matched controls across all conditions of the D-KEFS CWIT. Notably, the FLE group demonstrated larger impairment in the Inhibition and Inhibition/ Switching conditions. Additionally, a subgroup analysis indicated that patients with left-hemisphere FLE were the most disadvantaged in these conditions. However, the FLE and TLE groups did not significantly differ in their performance on the D-KEFS.

In another study, McDonald et al. (2005b) evaluated the D-KEFS TMT in patients with FLE and compared their performance with that of patients with TLE and healthy controls. Their results indicated that the FLE group performed disproportionately slower on the switching condition than the TLE and control groups, although the three groups were indistinguishable on the baseline conditions of the test. In terms of accuracy, the FLE group also committed more set-loss errors than the other two groups. Nevertheless, the performance of patients with TLE and healthy controls were comparable across all conditions, including the switching task. McDonald et al. (2005c) also found that patients with FLE generated significantly fewer accurate designs than TLE patients and healthy controls on the D-KEFS DFT switching condition. However, the three groups were otherwise comparable on the fundamental conditions of the test. Additionally, a subgroup analysis indicated that patients with left-sided FLE were the most disadvantaged on the switching task in terms of the number of correct designs generated, than those with right-sided FLE and controls. Those with left-lesional and non-lesional FLE were also found committing more set-loss errors on the switching task. Notably, when the cut-off for impairment was set at 1.5 standard deviations below the mean of the control group, the D-KEFS DFT

switching condition demonstrated a high specificity (90%) but only a modest sensitivity (57%) when discriminating between FLE and TLE patients. Lastly, another study by McDonald et al (2008) reported that patients with FLE performed significantly impaired than healthy controls on the D-KEFS PT but those with TLE did not differ from the controls on this task. When 1.5 standard deviation below control mean was defined as the impairment cut-off, the free inquiry condition demonstrated high specificity (80%) but a modest sensitivity (59%), whereas the multiple choice condition showed both comparatively lower sensitivity (55%) and specificity (75%). However, the discriminative ability of the free inquiry measure improved when only left-sided FLE patients were included.

DISCUSSION

Summary of evidence

This paper presents a systematic review of studies that evaluated evidence published in three databases between 2001 and 2018 with regard to the clinical utility of the D-KEFS for the identification of executive dysfunctions in clinical populations with acquired neurological conditions. In this part, the evidence from the studies was organised and discussed based on each of the research questions outlined in the objectives of the review.

D-KEFS performance in patients with various acquired neurological conditions

The majority of studies in this systematic review involved comparing the D-KEFS performance between participants with acquired neurological conditions and healthy individuals or between clinical groups. For studies related to traumatic brain injury and focal brain lesions, all of them found significant differences in D-KEFS performance between the clinical and control groups, with participants with TBI/ brain lesions performing significantly impaired when compared to

neurologically healthy controls on the selected D-KEFS measures. This conclusion is consistent across studies, with large to very small effect sizes being reported. Deficits in executive functioning as assessed by D-KEFS measures are also evident in other clinical groups. For instance, significant impairment in D-KEFS performance was reported in patients with FTD (Gansler et al., 2017; Huey et al., 2009; Kaiser et al., 2013); patients with MS (Parmenter et al., 2007) and people with MCI (Nutter-Upham et al., 2008) relative to healthy controls. For participants suffering from different types of epilepsy, findings suggest that both children (Luton et al., 2010; Parrish et al., 2007; Pulsipher et al., 2009) and adults (Reyes et al., 2017; McDonald et al., 2008; McDonald et al., 2005a; McDonald et al., 2005b; McDonald et al., 2005c) were significantly more impaired than healthy controls on D-KEFS tasks in terms of speed and accuracy. There is evidence to suggest that the clinical groups committed more errors than controls. For instance, patients with TBI/ brain lesions were reported to commit more set-switching and sequencing errors (Ghawami et al., 2017; Yochim et al., 2007); more errors on inhibition task (Faber et al., 2016); display a propensity to commit rule violations (Yochim et al., 2009); generate fewer correct responses on the D-KEFS VFT (Baldo et al., 2001); generate fewer correct responses and have to make more guesses on the D-KEFS WCT (Keil et al., 2005); as well as committing more rule violations in patients with FTD than healthy controls (Carey et al., 2008). These findings are consistent with known executive deficits of acquired neurological disorders.

Notably, there was one study investigating the cognitive impairment in people with ALS and PLS using D-KEFS and other cognitive measures, as well as examining how the deficits in cognitive performance were related to imaging metrics (Meoded et al., 2013). Although motor neurone disorders are most commonly characterised by progressive muscle wasting, weakness

and spasticity with relatively preserved cognition, much recent literature suggests that cognition dysfunctions are increasingly recognised in a subset of this patient group (Consonni et al., 2013; Oh et al., 2014; Rippon et al., 2006). Several neuropsychological studies have also suggested that mild to moderate deficits in executive functioning appear to represent a prominent feature of cognitive dysfunctions in people with ALS, with consistent findings being impaired verbal fluency, attention and memory function deficits (Consonni et al., 2013; Rippon et al., 2006; Zalonis et al., 2012). More importantly, among those ALS patients with cognitive impairment, about 20% to 30% of them manifest features of frontotemporal dementia (Consonni et al., 2013; Zalonis et al., 2012; Rippon et al., 2006). For these reasons, this study had been included in this review to question whether the D-KEFS can identify the specific cognitive profile in this client group. Despite the fact that study results showed no significant difference in D-KEFS performance between PLS and ALS patient groups, the D-KEFS "Fluency" and "Sorting" factors were linked with a reduced integrity of several white matter tracts connecting the frontal lobes with different brain regions. The findings appear to shed light on the view that executive dysfunctions can result from a disruption of the integrated network of connections in the brain system (Anderson, 2008), thereby supporting the clinical validity of the D-KEFS in its assessment of executive functioning.

As reported in the D-KEFS technical manual, the increased processing demands of the switching paradigms can provide additional clinical utility in the detection of subtle executive dysfunctions over traditional counterparts (Delis, Kaplan, & Kramer, 2001b). There is some evidence to suggest that patients with TBI (Anderson et al., 2017; Ghawami et al., 2017; Strong et al., 2011; Faber et al., 2016), brain lesions (Yochim et al., 2007) and epilepsy (Luton et al., 2010; McDonald et al., 2005a; McDonald et al., 2005b; McDonald et al., 2005c) were

disproportionately impacted by the switching conditions. Notably, the CWIT Inhibition/ Switching condition demonstrated a good predictive ability in identifying subsequent cognitive decline in a group of normal functioning older adults (Clark et al., 2012; Fine et al., 2008). However, the findings were inconsistent across studies. Two studies have shown that the clinical and control groups performed similarly on switching conditions (Keifer & Tranel, 2013; Baldo et al., 2001), suggesting that D-KEFS switching paradigms may not in particular a sensitive measure for executive dysfunctions. Finally, it is worth emphasizing that the generalisability of results should, however, be interpreted in light of the quality of studies.

Diagnostic accuracy of D-KEFS

Among the studies, only ten had documented the diagnostic accuracy metrics of D-KEFS subtests. Findings from those studies indicate that certain D-KEFS measures appear to have some value in distinguishing neurological pathologies from healthy controls. For instance, the D-KEFS ST "attempted sorts" was reported to be a more sensitive measure to distinguish group difference between people with severe TBI and healthy controls when compared to WCST, with 9.5 sorts as the optimal cut-off point to give an AUC of 0.8 (i.e. sensitivity = 88%; specificity = 65%) (Heled et al., 2012); the measure of committing two or more rule violations also demonstrated a good capability (sensitivity = 83%; specificity = 100%) for identifying people with frontal lesions from normal individuals (Yochim et al., 2009) as well as the "Switching" factor of the VFT was found to differentiate people with MCI from controls, with overall classification accuracy of 75% (Nutter-Upham et al., 2008). Nevertheless, there is also suggestive evidence that the D-KEFS showed some limitations to differentiate subjects with the disease condition between different clinical groups. For instance, Gansler et al. (2017) reported that the D-KEFS alone did not perform best in discriminating bvFTD from CBS and PPA.

Anderson et al. (2017) found that the combined CWIT Inhibition/ Switching and VFT Category Switching tasks lacked sufficient discriminative power to correctly identify people with moderate-to-severe TBI from the mild-uncomplicated TBI and healthy groups although the model performed well in identifying the presence of cognitive impairment in participants with TBI. Furthermore, McDonald et al. (2008) and McDonald et al. (2005c) reported that the PT and DFT switching condition only demonstrated modest sensitivities when discriminating between FLE and TLE patients. Carey et al. (2008) also found that the number of rule violations committed showed a low sensitivity of 50% when distinguishing FTD from AD.

Correlations of D-KEFS performance with brain regions

In this review, findings from the studies indicate that the executive dysfunction measured by the D-KEFS performance was mostly localised in the frontal brain regions (Gansler et al., 2017; Huey et al., 2009). For instance, verbal fluency was associated with areas of left frontal perisylvian cortex, sorting was related to dorsolateral prefrontal cortex and performance on TQT was correlated with left anterior frontal cortex (Huey et al., 2009). The results provide evidence for the role of prefrontal cortex in executive functioning. Apart from this, impaired D-KEFS scores were also correlated with reduced integrity of network connecting the frontal lobes with different regions of the brain (Meoded et al., 2013; Reyes et al., 2017); or with the frontal brain volumes (Carey et al., 2008; Fine et al., 2009; Kramer et al., 2007; Pulsipher et al., 2009); or with decreased structural integrity of brain tissues (Faber et al., 2016), which was reported as a consequence of the early TBI. Taken all together, the findings seem to lend some validity to the emerging view that executive functions are mediated by a widespread, integrated network of connections between the frontal lobes and other related subcortical structures throughout the brain (Anderson, 2008). From these findings, it can be seen that not only do certain D-KEFS

subtests show their utilities to the detection of executive impairments in frontal lobe injury, the D-KEFS also appears sensitive to other brain region damage. The neuroimaging findings therefore offer support for the construct validity of the D-KEFS, in particular its ability to tap into executive functioning and identify deficits in executive functions associated with the prefrontal cortex and its related circuitry.

Limitations

Finally, some limitations of this review should be considered. Firstly, there were a number of differences between the reviewed studies including the heterogeneous range of injuries and severities across the samples, the country and clinical settings in which participants were recruited from, the experimental paradigm used and how healthy controls were matched to the clinical groups. Hence, it is not possible to statistically pool data from each individual study, thereby precluding a comprehensive synthesis of results in a meta-analysis to measure the overall clinical utility of the D-KEFS. Secondly, the differences in clinical severity can influence the sensitivity and specificity of a test, which in turn affects its diagnostic accuracy (Wong & Lim, 2011). In this sense, studies may not be comparable if groups of subjects with significant differences in disease characteristics are compared. Because of this limitation a descriptive analysis was used in this review. Moreover, given that the studies evaluating the diagnostic accuracy of the D-KEFS are scarce, the results should be interpreted with caution. More studies involving diagnostic accuracy of the D-KEFS may be required for future research. Furthermore, since this review is based on literature search limited to articles published in three major databases, articles which are not published could not have been included through the search strategy. It is therefore possible that some studies have been missed and publication bias could be resulted. Egger, Jüni, Bartlett, Holenstein & Sterne (2003), however, have raised their

concern over the possibility of introducing bias when such literature is included, because these studies are often of lower methodological quality.

Clinical implications

The findings from this review appear to indicate that the D-KEFS is clinically useful in several ways. Firstly, available evidence does suggest that the various D-KEFS subtests demonstrate reasonable sensitivity in distinguishing many different types of acquired neurological conditions from healthy controls. It has been known that the selection of the D-KEFS subtests varied across the reviewed studies. Study results have shown that either various individual subtests or combinations of subtests were useful in assessing executive dysfunctions in people with neurological pathologies, including children and adults. Clinicians can therefore be flexible in terms of administering specific part of the D-KEFS, depending on the specific needs of the client or the time constraints of the clinician. This makes the D-KEFS a good choice of instrument for the assessment of executive functioning. Secondly, there are various suggestive pieces of evidence that people with acquired neurological conditions committed more perseverative or set-loss errors than healthy controls. The application of the additional error analysis thus allows clinicians to gain more information regarding each client's specific profile of executive functions, thereby offering opportunities for clinicians in a multi-dimensional approach to assessment. Furthermore, this systematic review also showed that the D-KEFS performance was correlated with some brain regions. This appears to confirm the construct validity of the D-KEFS which assists clinicians in the refinements to this complex construct.

Taken all together, the D-KEFS is concluded to be a promising assessment tool in clinical settings. However, clinicians should also be mindful that the diagnostic accuracy of the test can

change depending on the type and nature of the neurological pathology assessed, disease severity or what D-KEFS subtests or conditions are used. As evidenced from the findings, for instance, it is expected that the test has greater sensitivity when it is used in distinguishing FTD from healthy population than from AD. Therefore, clinicians should be cautious not to over-interpret performances on individual D-KEFS subtests. Rather, it is advisable to integrate test scores with behavioural observations, background information and other qualitative aspects of performance during testing.

CONCLUSIONS

This review suggests that the D-KEFS appears to be a useful evaluation tool of executive functioning, based on the available evidence. Its findings indicate that participants with various acquired neurological conditions showed significant executive impairment, including committing more errors on the D-KEFS than healthy individuals. The performance on the D-KEFS was also correlated with frontal lobes and other brain regions, which qualifies the D-KEFS as an assessment tool for identifying executive dysfunctions in different clinical populations associated with the prefrontal cortex. Moreover, the D-KEFS may have some value to discriminate neurological pathologies from within healthy population, although the evidence available is insufficient. Furthermore, whether the D-KEFS switching paradigm is sensitive to subtle executive function deficits is still in doubt unless future research is conducted which proves otherwise. Whilst some studies in this review had potential risk of bias due to the relatively small sample sizes, generally the quality of the studies was deemed sufficient for meaningful synthesis and analysis. In summary, this review highlighted that the D-KEFS is clinically useful despite the limitations mentioned above.

REFERENCES

Alvarez, J., & Emory, E. (2006). Executive Function and the Frontal Lobes: A meta-analytic review. *Neuropsychology Review*, *16*(1), 17-42. doi:10.1007/s11065-006-9002-x

Anderson, P. (2008). Towards a developmental model of executive function. In V. Anderson, R. Jacobs, & P. Anderson (Eds.), *Executive functions and the frontal lobes: A lifespan perspective* (pp. 3–21). New York, NY: Taylor and Francis.

Anderson, L. B., Jaroh, R., Smith, H., Strong, C. A. H., & Donders, J. (2017). Criterion validity of the D-KEFS colour–word and verbal fluency switching paradigms following traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *39*(9), 890-899. doi:10.1080/13803395.2016.1277513

Baldo, J. V., Delis, D. C., Wilkins, D. P., & Shimamura, A. P. (2004). Is it bigger than a breadbox? Performance of patients with prefrontal lesions on a new executive function test. *Archives of Clinical Neuropsychology*, *19*(3), 407-419. doi:10.1016/S0887-6177(03)00074-X

Baldo, J. V., Shimamura, A. P., Delis, D. C., Kramer, J., & Kaplan, E. (2001). Verbal and design fluency in patients with frontal lobe lesions. *Journal of the International Neuropsychological Society*, *7*(5), 586-596.

Baron, I. S. (2004). Delis-Kaplan Executive Function System. *Child Neuropsychology*, *10*(2), 147-152. doi:10.1080/09297040490911140

Carey, C. L., Woods, S. P., Damon, J., Halabi, C., Dean, D., Delis, D. C., ... Kramer, J. H. (2008). Discriminant validity and neuroanatomical correlates of rule monitoring in frontotemporal dementia and Alzheimer's disease. *Neuropsychologia*, *46*(4), 1081-1087. doi:10.1016/j.neuropsychologia.2007.11.001

Chan, R., Shum, D., Toulopoulou, T., & Chen, E. (2008). Assessment of executive functions:

Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201-216. doi:10.1016/j.acn.2007.08.010

Chesters, M. J. (2008). D-KEFS validity: An update of the research. Technical Report. Oxford: Pearson Assessment.

Cicerone, K. D. (2005). Rehabilitation of executive impairments. In W. M. High, A. M. Sander, M. A. Struchen, & K. A. Hart (Eds.), *Rehabilitation for traumatic brain injury* (2nd ed., pp. 12-21). New York, NY: Oxford University Press.

Clark, L. R., Schiehser, D. M., Weissberger, G. H., Salmon, D. P., Delis, D. C., & Bondi, M. W. (2012). Specific measures of executive function predict cognitive decline in older adults. *Journal of the International Neuropsychological Society*, *18*(1), 118-127. doi:10.1017/S1355617711001524.

Consonni, M., Iannaccone, S., Cerami, C., Frasson, P., Lacerenza, M., Lunetta, C., … Cappa, S. F. (2013). The cognitive and behavioural profile of amyotrophic lateral sclerosis: Application of the consensus criteria. Behavioural *Neurology*, *27*(2), 143-153. doi:10.3233/BEN-2012-110202

Coughlan, M., Cronin, P., & Ryan, F. (2007). Step-by-step guide to critiquing research. Part 1: quantitative research. *British Journal of Nursing*, *16*(11), 658-663.

Crawford, J. (1998). Introduction to the assessment of attention and executive functioning. *Neuropsychological Rehabilitation*, *8*(3), 209-211. doi:10.1080/713755574

Cummings, J. L., & Miller, B. L. (2007). Conceptual and clinical aspects of the frontal lobes. In B. L. Miller, & J. L. Cummings (Eds.), *The human frontal lobes: functions and disorders* (2nd ed., pp. 12-21). New York, NY: Guilford Press.

Delis, D. C., Kaplan, E., & Kramer, J. (2001a). *Delis-Kaplan Executive Function System: Examiner's manual.* San Antonio, TX: Psychological Corporation.

Delis, D. C., Kaplan, E., & Kramer, J. (2001b). *Delis-Kaplan Executive Function System: Technical manual.* San Antonio, TX: Psychological Corporation.

Delis, D.C., Kramer, J., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan Executive Function System: an update. *Journal of the International Neuropsychological Society*, *10*(2), 301-303. doi:10.1017/s1355617704102191

Dubois, B., Slachevsky, A., Litvan, I., & Pillon, B. (2000). The FAB: A frontal assessment battery at bedside. *Neurology*, *55*(11), 1621-1626. doi:10.1212/wnl.55.11.1621

Egger, M., Juni, P., Bartlett, C., Holenstein, F., & Sterne, J. (2003). How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment*, *7*(1), 1-76.

Faber, J., Wilde, E. A., Hanten, G., Ewing-Cobbs, L., Aitken, M. E., Yallampalli, R., ... Hunter, J. V. (2016). Ten-year outcome of early childhood traumatic brain injury: Diffusion tensor imaging of the ventral striatum in relation to executive functioning. *Brain injury*, *30*(13-14), 1635-1641. doi:10.1080/02699052.2016.1199910

Fine, E. M., Delis, D. C., Dean, D., Beckman, V., Miller, B. L., Rosen, H. J., & Kramer, J. H. (2009). Left frontal lobe contributions to concept formation: A quantitative MRI study of performance on the Delis–Kaplan Executive Function System Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, *31*(5), 624-631. doi:10.1080/13803390802419017

Fine, E. M., Delis, D. C., Wetter, S. R., Jacobson, M. W., Jak, A. J., McDonald, C. R., ... Bondi, M. W. (2008). Cognitive discrepancies versus APOE genotype as predictors of cognitive decline in normal-functioning elderly individuals: a longitudinal study. *The American Journal of Geriatric Psychiatry*, *16*(5), 366-374.

Gansler, D. A., Huey, E. D., Pan, J. J., Wasserman, E., & Grafman, J. H. (2017). Assessing

the dysexecutive syndrome in dementia. *J Neurol Neurosurg Psychiatry*, *88*(3), 254-261. doi:10.1136/jnnp-2016-313576

Ghawami, H., Sadeghi, S., Raghibi, M., & Rahimi-Movaghar, V. (2017). Executive functioning of complicated-mild to moderate traumatic brain injury patients with frontal contusions. *Applied Neuropsychology: Adult*, *24*(4), 299-307. doi:10.1080/23279095.2016.1157078

Goldberg, E., & Bougakov, D. (2005). Neuropsychologic assessment of frontal lobe dysfunction. *Psychiatric Clinics of North America*, *28*(3), 567-580. doi:10.1016/j.psc.2005.05.005

Heled, E., Hoofien, D., Margalit, D., Natovich, R., & Agranov, E. (2012). The Delis–Kaplan Executive Function System Sorting Test as an evaluative tool for executive functions after severe traumatic brain injury: A comparative study. *Journal of Clinical and Experimental Neuropsychology*, *34*(2), 151-159. doi:10.1080/13803395.2011.625351

Homack, S., Lee, D., & Riccio, C. (2005). Test Review: Delis-Kaplan Executive Function System. *Journal of Clinical and Experimental Neuropsychology*, *27*(5), 599-609. doi:10.1080/13803390490918444

Houston, W. S., Delis, D. C., Lansing, A., Jacobson, M. W., Cobell, K. R., Salmon, D. P., & Bondi, M. W. (2005). Executive function asymmetry in older adults genetically at-risk for Alzheimer's disease: Verbal versus design fluency. *Journal of the International Neuropsychological Society*, *11*(7), 863-870. doi:10.10170S1355617705051015

Huey, E. D., Goveia, E. N., Paviol, S., Pardini, M., Krueger, F., Zamboni, G., ... Grafman, J. (2009). Executive dysfunction in frontotemporal dementia and corticobasal syndrome. *Neurology*, *72*(5), 453-459.

Irwig, L., Bossuyt, P., Glasziou, P., Gatsonis, C., & Lijmer, J. (2002). Designing studies to

ensure that estimates of test accuracy are transferable. *British Medical Journal*, *324*, 669-671. doi:10.1136/bmj.324.7338.669

Jurado, M., & Rosselli, M. (2007). The Elusive Nature of Executive Functions: A review of our current understanding. *Neuropsychology Review*, *17*(3), 213-233. doi:10.1007/s11065-007-9040-z

Kaiser, N. C., Lee, G. J., Lu, P. H., Mather, M. J., Shapira, J., Jimenez, E., ... Mendez, M. F. (2013). What dementia reveals about proverb interpretation and its neuroanatomical correlates. *Neuropsychologia*, *51*(9), 1726-1733. doi.org/10.1016/j.neuropsychologia.2013.05.021

Keifer, E., & Tranel, D. (2013). A neuropsychological investigation of the Delis-Kaplan executive function system. *Journal of Clinical and Experimental Neuropsychology*, *35*(10), 1048-1059. doi:10.1080/13803395.2013.854319

Keil, K., Baldo, J., Kaplan, E., Kramer, J., & Delis, D. C. (2005). Role of frontal cortex in inferential reasoning: evidence from the word context test. *Journal of the International Neuropsychological Society*, *11*(4), 426-433. doi:10.10170S1355617705050551

Kramer, J. H., Quitania, L., Dean, D., Neuhaus, J., Rosen, H. J., Halabi, C., ... Miller, B. L. (2007). Magnetic resonance imaging correlates of set shifting. *Journal of the International Neuropsychological Society*, *13*(3), 386-392.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th ed.). New York, NY: Oxford University Press.

Luton, L. M., Burns, T. G., & DeFilippis, N. (2010). Frontal lobe epilepsy in children and adolescents: a preliminary neuropsychological assessment of executive function. *Archives of Clinical Neuropsychology*, *25*(8), 762-770.

McDonald, C. R., Delis, D. C., Kramer, J. H., Tecoma, E. S., & Iragui-Madoz, V. J. (2008).

A componential analysis of proverb interpretation in patients with frontal lobe epilepsy and temporal lobe epilepsy: Relationships with disease-related factors. *The Clinical Neuropsychologist*, *22*(3), 480-496. doi:10.1080/13854040701363828

McDonald, C. R., Delis, D. C., Norman, M. A., Tecoma, E. S., & Iragui-Madoz, V. J. (2005b). Is impairment in set-shifting specific to frontal-lobe dysfunction? Evidence from patients with frontal-lobe or temporal-lobe epilepsy. *Journal of the International Neuropsychological Society*, *11*(4), 477-481. doi:10.10170S1355617705050484

McDonald, C. R., Delis, D. C., Norman, M. A., Tecoma, E. S., & Iragui-Madoz, V. J. (2005c). Discriminating patients with frontal-lobe epilepsy and temporal-lobe epilepsy: utility of a multilevel design fluency test. *Neuropsychology*, *19*(6), 806. doi:10.1037/0894-4105.19.6.806

McDonald, C. R., Delis, D. C., Norman, M. A., Wetter, S. R., Tecoma, E. S., & Iragui-Madoz, V. J. (2005a). Response inhibition and set shifting in patients with frontal lobe epilepsy or temporal lobe epilepsy. *Epilepsy & Behaviour*, *7*(3), 438-446. doi:10.1016/j.yebeh.2005.05.005

Meoded, A., Kwan, J. Y., Peters, T. L., Huey, E. D., Danielian, L. E., Wiggs, E., ... Grafman, J. (2013). Imaging findings associated with cognitive performance in primary lateral sclerosis and amyotrophic lateral sclerosis. *Dementia and Geriatric Gognitive Disorders Extra*, *3*(1), 233-250. doi:10.1159/000353456

Miyake, A., Emerson, M., & Friedman, N. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language*, *Volume 21*(2), 0169-0183. doi:10.1055/s-2000-7563

Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49-100. doi:10.1006/cogp.1999.0734

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264-269. doi.org/10.1371/journal.pmed.1000097

Nutter-Upham, K. E., Saykin, A. J., Rabin, L. A., Roth, R. M., Wishart, H. A., Pare, N., & Flashman, L. A. (2008). Verbal fluency performance in amnestic MCI and older adults with cognitive complaints. *Archives of Clinical Neuropsychology*, *23*(3), 229-241. doi:10.1016/j.acn.2008.01.005

Oh, S. I., Park, A., Kim, H. J., Oh, K. W., Choi, H., Kwon, M. J., … Kim, S. H. (2014). Spectrum of cognitive impairment in Korean ALS patients without known genetic mutations. *PLoS ONE*, *9*(2), e87163. doi:10.1371/journal.pone.0087163

Parmenter, B. A., Zivadinov, R., Kerenyi, L., Gavett, R., Weinstock-Guttman, B., Dwyer, M. G., ... Benedict, R. H. (2007). Validity of the Wisconsin card sorting and Delis–Kaplan executive function system (DKEFS) sorting tests in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, *29*(2), 215-223. doi: 10.1080/13803390600672163

Parrish, J., Geary, E., Jones, J., Seth, R., Hermann, B., & Seidenberg, M. (2007). Executive functioning in childhood epilepsy: Parent-report and cognitive assessment. *Developmental Medicine & Child Neurology*, *49*(6), 412-416.

Pulsipher, D. T., Seidenberg, M., Guidotti, L., Tuchscherer, V. N., Morton, J., Sheth, R. D., & Hermann, B. (2009). Thalamofrontal circuitry and executive dysfunction in recent-onset juvenile myoclonic epilepsy. *Epilepsia*, *50*(5), 1210-1219. doi: 10.1111/j.1528-1167.2008.01952.x

Reyes, A., Uttarwar, V. S., Chang, Y. H. A., Balachandra, A. R., Pung, C. J., Hagler Jr, D. J., ... McDonald, C. R. (2018). Decreased neurite density within frontostriatal networks is associated with executive dysfunction in temporal lobe epilepsy. *Epilepsy & Behaviour*, *78*, 187-193. doi.org/10.1016/j.yebeh.2017.09.012

Rippon, G. A., Scarmeas, N., Gordon, P. H., Murphy, P. L., Albert, S. M., Mitsumoto, H., …
Stern, Y. (2006). An observational study of cognitive impairment in amyotrophic lateral
sclerosis. *Archives of Neurology*, *63*(3), 345-352.

Schmidt, M. (2003). Hit or miss? Insight into executive functions. *Journal of the
International Neuropsychological Society*, *9*(6), 962–964.
doi:10.1017/s1355617703230162

Shunk, A. W., Davis, A. S., & Dean, R. S. (2006). Test review: Delis-Kaplan Executive
Function System (D-KEFS). *Applied Neuropsychology*, *13*(4), 275-27.
doi:10.1207/s15324826an1304_9

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

Strong, C. A. H., Tiesma, D., & Donders, J. (2011). Criterion validity of the Delis-Kaplan
Executive Function System (D-KEFS) fluency subtests after traumatic brain
injury. *Journal of the International Neuropsychological Society*, *17*(2), 230-237.
doi:10.1017/S1355617710001451

Stuss, D., & Alexander, M. (2000). Executive functions and the frontal lobes: a conceptual
view. *Psychological Research*, *63*(3), 289-298. doi:10.1007/s004269900007

Suchy, Y. (2016). *Executive functioning: A comprehensive guide for clinical practice*. New
York, NY: Oxford University Press.

Swanson, J. (2005). The Delis-Kaplan executive function system: a review. *Canadian
Journal of School Psychology*, *20*(1-2), 117-128. doi:10.1177/0829573506295469

Testa, R., Bennett, P., & Ponsford, J. (2012). Factor analysis of nineteen executive function
tests in a healthy adult population. *Archives of Clinical Neuropsychology*, *27*(2), 213-
224. doi.org/10.1093/arclin/acr112

Vandenbroucke, J. P., Von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., ... Egger, M. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Epidemiology*, *18*(6), 805-835.

Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Annals of Internal Medicine*, *147*(8), 573-577.

Wetter, S. R., Delis, D. C., Houston, W. S., Jacobson, M. W., Lansing, A., Cobell, K., ... Bondi, M. W. (2005). Deficits in inhibition and flexibility are associated with the APOE-E4 allele in non-demented older adults. *Journal of Clinical and Experimental Neuropsychology*, *27*(8), 943-952. doi:10.1080/13803390490919001

Wong, H. B., & Lim, G. H. (2011). Measures of diagnostic accuracy: Sensitivity, specificity, PPV and NPV. *Proceedings of Singapore healthcare*, *20*(4), 316-318.

Yochim, B. P., Baldo, J. V., Kane, K. D., & Delis, D. C. (2009). D-KEFS Tower Test performance in patients with lateral prefrontal cortex lesions: the importance of error monitoring. *Journal of Clinical and Experimental Neuropsychology*, *31*(6), 658-663. doi:10.1080/13803390802448669

Yochim, B., Baldo, J., Nelson, A., & Delis, D. C. (2007). D-KEFS Trail Making Test performance in patients with lateral prefrontal cortex lesions. *Journal of the International Neuropsychological Society*, *13*(4), 704-709. doi:10.10170S1355617707070907

Zalonis, I., Christidi, F., Paraskevas, G., Zabelis, T., Evdokimidis, I., & Kararizou, E. (2012). Can executive cognitive measures differentiate between patients with spinal- and bulbar-onset amyotrophic lateral sclerosis? *Archives of Clinical Neuropsychology*, *27*(3), 348-354. doi.org/10.1093/arclin/acs031

EMPIRICAL PAPER:

INVESTIGATING THE VALIDITY OF THE DELIS-KAPLAN EXECUTIVE
FUNCTION SYSTEM (D-KEFS) AS A NEUROPSYCHOLOGICAL ASSESSMENT
TOOL FOR EXECUTIVE FUNCTIONS IN THE TRAUMATIC BRAIN INHURY (TBI)
IN THE UK

by

Yin-Ming Chan

ABSTRACT

Background

Deficits in executive functioning are highly prevalent in people with Traumatic Brain Injury (TBI). The Delis Kaplan Executive Function (D-KEFS) comprises a standardised set of tests designed to measure a wide spectrum of abilities associated with executive functioning. Currently there is substantial evidence to support the validity of D-KEFS as a useful instrument to identify deficits in executive functioning. This study aims to investigate the validity of the D-KEFS by comparing the performance of a sample of patients with mild-uncomplicated to severe TBI, with that of orthopaedic controls using selected D-KEFS subtests. The orthopaedic patients are considered as 'gold standard' controls for studying TBI given they are arguably representative of the TBI group both demographically and psychosocially.

Methods

One hundred patients with mild-uncomplicated to severe Traumatic Brain Injury (TBI) and twenty-six orthopaedic patients were recruited. Measures of performance validity were administered to participants. Those who failed the tests were excluded from the study. Selected D-KEFS subtests (Trail Making, Verbal Fluency, Colour Word Interference Test and Tower Test) were administered to both TBI and orthopaedic groups, as well as the application of additional measures of premorbid intellectual functioning and mental processing speed to account for other influences on D-KEFS performance that were not specific to executive functioning.

Results

The TBI group performed significantly worse than the orthopaedic controls on the Trail Making Number-Letter Switching, Colour Word Interference Inhibition, Colour Word Interference Inhibition/ Switching, Letter Fluency and Category Switching tasks, but not on Category Fluency and Tower Test. The Executive Functioning Indices (EFIs) constructed also demonstrated significant group mean differences, with TBI patients performing worse than the orthopaedic controls. More importantly the composite EFIs resulted in greater effect sizes than the individual subtests. Notably, group differences vanished when the effect of processing speed was controlled for. Lastly, the results also indicated that the TBI group presented with a consistently higher rate of obtaining low scores in comparison to the orthopaedic controls.

Conclusions

To our knowledge this is the first study in the UK that has compared the performance of a sample of patients with mild-uncomplicated to severe TBI, with that of orthopaedic controls using D-KEFS. Relative to the orthopaedic controls the TBI patients showed marked deficits in executive functioning across some subtests as well as in EFIs, with moderate to large effect sizes. These findings support the use of the D-KEFS EFIs for the assessment of executive functioning in TBI population. Future study might consider using a larger and more gender balanced orthopaedic population and also examine the performances of TBI individuals suffering head trauma of various aetiologies on D-KEFS in order to capture a more comprehensive cognitive profile for this population group.

INTRODUCTION

Traumatic brain injury (TBI) is a global public health issue worldwide (World Health

Organisation, 2006). In the UK, it is estimated that around 1.4 million people attended hospitals

for head-related injury each year, causing significant mortality and disability in people under the

age of 40 years (National Clinical Guideline Centre, 2014). Epidemiological data from a study

conducted at a UK hospital indicated that those who presented to the emergency department for

head-related injury accounted for 3.4% of all annual attendances (Yates, Williams, Harris,

Round & Jenkins, 2006). Moreover, of these, 10.9% were classified as moderate to severe cases.

Furthermore, the highest rates of admission for moderate to severe head injury were recorded

amongst adolescent males aged between 15−19 years, although it is still worth noting that

children under five years of age living in urban areas are also reported to have a higher risk of

sustaining head injury. In another survey done by the Headway Brain Injury Association, all

admission statistics related to acquired brain injury in the UK hospitals were compiled, showing

that a total of 155,919 hospital admissions for head injury was reported in 2016-17 ("Acquired

brain injury 2016-2017 statistics based on UK admissions," n.d.). Additionally, men were 1.5

times more likely to be admitted than women but female head injury admissions have

significantly risen by 23% since 2005-6 ("Acquired brain injury 2016-2017 statistics based on

UK admissions," n.d.). According to the Department of Health (2005), it has been estimated that

there are as many as 500,000 people aged between 16 and 74 years across the UK currently

living with long-term disabilities as a consequence of TBI.

Significantly, those who sustain moderate to severe TBI often experience cognitive, emotional

and behavioural sequelae that have an adverse impact on their quality of life (Corrigan &

Hammond, 2013). Among the long-term outcomes following TBI, neurocognitive sequelae are

cited as a major cause of disability which can affect the injured individuals' everyday life
activities as well as that of their families (Fleminger & Ponsford, 2005; Rabinowitz & Levin,
2014). Common cognitive sequelae resulting from injury include difficulties in attention,
memory, information-processing speed and executive functions (Iverson, Holdnack & Lange,
2013). However, impairment can be variable ranging from mild to severe, often depending on
the nature and severity of the injury (Rabinowitz & Levin, 2014).

TBI is an acquired brain injury (ABI) caused by the application of an external mechanical force
to the head. It is recognised that frontal lobes, temporal lobes and related circuitry such as
subcortical white matter, basal ganglia and thalamus are particularly vulnerable to TBI due to
their locations and the large areas involved (Smith, 2011). Diffuse axonal injury (DAI) is also
prevalent across all levels of injury severity, resulting in a disruption of the integrated network of
connections between the frontal lobes and various subcortical structures (Smith, 2011). All of
these factors explain why deficits in executive functioning are highly prevalent in people with
TBI (Suchy, 2016).

Historically, the frontal lobes have been conceptualised as the main seat of higher-order
cognitive processes that play a large role in executive functioning (Tranel, Anderson & Benton,
1994; Fuster, 1993). This association is largely driven by early observations and case studies
involving frontal brain damage (Luria, 1966, 1973; Stuss & Alexander, 2000, 2007). However,
this understanding may not be entirely accurate (Alvarez & Emory, 2006). More recently, with
the advancement of neuroimaging techniques, studies have provided support to the theory that
executive function is not simply limited to the frontal region as hypothesized previously in the

literature. Instead, various integrated networks and subcortical structures appear to be also involved in executive functioning (Chung, Weyandt & Swentosky, 2014).

Whilst there have been many definitions proposed for executive function, Goldstein, Naglieri, Princiotta, & Otero (2014) have suggested that executive function is a multifaceted construct that involves a variety of high-level cognitive processes mediated by prefrontal areas of the frontal lobes, including but not limited to planning, attention, working memory, initiation, self-monitoring for goal progress and errors, self-regulation and inhibition, as well as to produce goal-oriented behaviours (Jurado & Rosselli, 2007). More specifically, Suchy (2016) has defined executive functioning as "an umbrella term that subsumes a set of higher-order top-down neurocognitive processes involved in planning, selection and execution of actions that are purposeful and adaptive, goal-directed and future-orientated, and socially informed" (Cummings & Miller, 2007; Gazzaley & D'Esposito, Miller & Cummings, 2007; Lezak, Howieson, Bigler, Tranel, 2013 cited in Suchy, 2016, p 10). Suchy (2016) has further conceptualised the construct of executive function using five subdomains, based on its purposes from an evolutionary perspective (as well as the associated syndromes that result from the specific deficit). These include executive cognitive functions (dysexecutive syndrome), meta-tasking (disorganised manner), response selection (disinhibition syndrome), initiation/ maintenance (apathetic syndrome) and social cognition (inappropriate syndrome). This definition provides a useful structured framework for the development of more targeted assessments of executive functions making distinctions possible between these subdomains.

At present there are a range of cognitive measures pertaining to the evaluation of executive functions. Some common traditional tests of executive functioning include the Wisconsin Card

Sorting Test (WCST), Trail Making Test (TMT), Tower of London tasks and measures of fluency. Although these measures are widely used in both clinical and research settings, reliance on them solely for the measurement of executive function is not without its psychometric challenges. Firstly, it is recognised that measures of executive functioning tend to have relatively low internal and test-retest reliability (Rabbitt, 1997), meaning that measurements are not consistent over time and that the test scores can potentially lead to misleading results. For instance, according to a recent meta-analysis, the test-retest reliability of WCST Perseverative Errors was lower (i.e. Pearson's r <0.7) when compared to other neuropsychological measures such as Wechsler Adult Intelligence Scale (Calamia, Markon & Tranel, 2013). Suchy (2016) has also suggested that some tests of executive functioning (e.g. WCST) do not readily lend themselves to this type of reliability measurement because the construct being measured will be fundamentally altered in a repeated administration. As a consequence this can lead to a constriction of range in test scores when the test is administered more than once, thereby lowering the test-retest correlations.

Delis, Kramer, Kaplan & Holdnack (2004) have also explained that executive functioning tests tap into wider complex cognitive processes rather than a more homogeneous and fundamental construct due to the multifaceted nature of executive functioning. This makes measures of executive functioning more susceptible to performance variability, limiting reliability values. Moreover, in addition to issues of reliability, many executive functioning tasks are also found to have very low ecological validity. This means that patients' performances on these tasks do not necessarily correlate well with their functioning in daily life (Barkley, 2011; Ardila, 2008). Furthermore, other scholars have also argued that the construct validities of executive functioning tasks are often not well-established (Miyake, Emerson & Friedman, 2000) despite

the fact that they have been used for many years. Given the variability in the way that the construct of executive functioning is defined and conceptualised, establishing construct validity for these measures is thus not easy to accomplish (Suchy, 2016).

The Delis Kaplan Executive Function (D-KEFS) is a standardised set of tests designed to measure a wide spectrum of abilities associated with executive functioning. Originally published in 2001, this test is composed of nine stand-alone subtests that evaluate higher-order cognitive functions in both children and adults from 8 to 89 years of age. Each subtest can be either individually administered or used in combination with other subtests, depending on specific assessment needs. The individual subtest is largely derived from long-standing or previously used neuropsychological tests (e.g. Stroop Colour Word Test, Trail Making Test, Tower of Hanoi, California Card Sorting Test).

Notably, one of the strengths of the D-KEFS that differentiate it from other tests of executive functioning is the use of a cognitive-process approach in the interpretation of test scores, which provides information in terms of identifying any neurocognitive mechanisms underlying poor performance under different test conditions (Homack, Lee & Riccio, 2005; Swanson, 2005). In addition, this battery of tests also include error measures which provides important qualitative information during interpretation. Several new switching conditions were also added to some of the D-KEFS subtests, making it more advantageous over traditional measures for the detection of subtle deficits in executive functioning (Swanson, 2005). According to a recent survey of neuropsychological assessment practice amongst members of the International Neuropsychological Society and the National Academy of Neuropsychology (LaDuke, Barr, Brodale & Rabin, 2018), the D-KEFS as a full battery was listed as the third most commonly

utilised test for the assessment of executive functioning, whereas the first two were the WCST and non-DKEFS TMT respectively. In addition to this, a number of individual D-KEFS subtests were also cited as being used specifically by some of the respondents, including the Colour Word interference test (10%), Verbal Fluency (8%), Tower Test (5%), Design Fluency, Trail Making and Twenty Questions Test (all used by 3% of respondents).

There is substantial evidence to support the validity of D-KEFS as a useful instrument to identify deficits in executive functioning. For instance, patients with dorsolateral prefrontal cortex brain lesions performed significantly more impaired than patients with ventromedial prefrontal cortex and non-frontal lesions on a number of D-KEFS measures (Keifer & Tranel, 2013). In another study, researchers found that patients with lateral prefrontal cortex lesions performed significantly worse than healthy controls on the D-KEFS Tower test (Yochim, Baldo, Kane & Delis, 2009). A similar pattern was also found for the D-KEFS Word Context Test with frontal lesion patients performing worse than controls (Keil, Baldo, Kaplan, Kramer & Delis, 2005).

Moreover, validity studies investigating the executive functions in people with TBI also indicate that various D-KEFS subtests demonstrate reasonable criterion validity in discriminating TBI patients from normal controls. For instance, Anderson, Jaroh, Smith, Strong & Donders (2017) found that performances on switching conditions are sensitive to particular TBI-severity indicators such as length of coma and nature of lesions. Longer length of coma was also related to the poor performance on the Colour Word Inhibition/ Switching task, whereas diffuse lesions predicted worse performance in Category Switching. Heled, Hoofien, Margalit, Natovich & Agranov (2012) also reported that the D-KEFS Sorting Test was more sensitive than WCST in distinguishing group difference on certain measures. Whilst these studies provide some evidence

on the clinical utility and usefulness of the D-KEFS in TBI, there are nevertheless only a few and most have relatively modest sample sizes.

With regard to the reliability of the D-KEFS, it was estimated using alternate form reliability and test-retest reliability (Delis, Kaplan & Kramer, 2001a). Depending on the particular measure used as well as the age group, the reliability values of the D-KEFS showed variability, ranging from low ($\leq$0.59) to high (0.8-0.89) (Strauss, Sherman & Spreen, 2006). Yet, many of the D-KEFS scores have been reported to have low reliability values (Crawford, Sutherland & Garthwaite, 2008; Schmidt, 2003). To maximise reliability, Crawford, Garthwaite, Sutherland & Borland (2011) and Suchy (2016) have suggested the construction of a D-KEFS composite Executive Functioning Index (EFI) score from the individual subtest scores. Suchy (2016) has cited a number of studies in which such an approach was adopted and in many cases the composite scores have reliabilities above 0.7. As an example, in a study examining executive functioning in older adults (Puente, Lindbergh & Miller, 2015), the reliabilities for the four D-KEFS subtests used (Trail Making condition 4, Letter Fluency, Design Fluency all 3 conditions and Tower test) range from 0.50 to 0.70. The resulted composite D-KEFS score, however, had a reliability of 0.75. As such, the process of developing a composite score is shown to be more reliable than using individual subtests. Given the insufficient reliability for most measures, the administration of multiple measures is seen to be advisable (Suchy, 2016). Additionally, this approach confers the added advantage of examining variability across scores and also allowing multivariate interpretation, thereby minimising the risk of attributing a small number of low scores as being indicative of impairment when in fact they may represent normal performance variations (Binder, Iverson & Brooks, 2009).

This present study thus aims to extend the validation data of the D-KEFS in people with TBI, using a larger sample than has been reported to date. Acute orthopaedic patients with no sustained head injury were included as controls in the current study instead of a normal population. This distinct clinical population is considered to be a 'gold standard' control for studying TBI given they are arguably representative of the TBI group both demographically and psychosocially (McKinlay & Brooks, 1984). As such, some confounding pre- or post-injury variables such as low education, lifestyle and the effects of general trauma which can affect neuropsychological performance can be better accounted for (McKinlay & Brooks, 1984).

The first objective of this study was to compare the performance of a sample of patients with mild-uncomplicated to severe TBI, with that of orthopaedic controls using selected D-KEFS subtests. Based on evidence from previous literature, it was hypothesised that people with TBI would show significant executive impairments, performing worse than controls on the measures. The second objective was to construct a D-KEFS Executive Functioning Index (EFI) as outlined above and investigate its utility to TBI in terms of its ability to discriminate between TBI group and orthopaedic controls. Specifically, the validity of the EFIs was further investigated by assessing any relationship between TBI severity and performance on EFIs.

METHODS

Participants

One hundred patients with Traumatic Brain Injury (TBI) were recruited from a consecutive cohort of patients seen in the Traumatic Brain Injury outpatient neuropsychology clinic at the major trauma unit of University Hospital Birmingham, UK. Although the majority of recovery from TBI takes place within the first two years after injury, some patients may still exhibit further recovery up to 3 to 5 years following severe injury (Rabinowitz & Levin, 2014). In order to ensure all participants were neuro-psychologically stable and in the chronic stage of recovery, patients were considered appropriate to participate in the study if they had experienced brain injury within the last three years. Patients were excluded from the study if they did not have sufficient mental capacity to be able to give informed consent for their participation, as judged by their treating clinician. Another selection criteria was that all the participants in this study were administered a Performance Validity Test (PVT) to measure their effort and motivation. Patient who did not pass the PVT were excluded. Given that the validity of neuropsychological findings fundamentally relies on the assumption that examinees have to perform the tests at their best level of ability, neuropsychologists commonly use PVT as part of their assessment. Consistent with PVTs being reported to be insensitive to the effects of TBI on memory but highly sensitive to effort (TOMM; Tombaugh, 1997; Green's WMT; Green, Allen & Astner, 1995), the distribution of failures did not appear to demonstrate a relationship with TBI severity: Mild-uncomplicated failures N = 1 (10%); Mild-complicated failures N = 3 (30%); Moderate TBI failures N = 4 (40%); and Severe TBI failures N = 2 (20%). The importance of excluding the PVT failures was also illustrated by the occurrence of consistently lower D-KEFS scores in those failing the PVT as opposed to those passing the test. The differences in performance on D-KEFS component scores between PVT passers and failers are shown in Table 5.

Table 5: Performance difference on D-KEFS component scores between PVT passers and failers

D-KEFS component scores	PVT Fail			PVT Pass		
	Mean	SD	N	Mean	SD	N
Trail Making Number-Letter Switching	***5.00a***	***3.59***	***10***	***8.43b***	***3.65***	***89***
Colour Word Interference Inhibition	6.75a	2.96	8	9.07a	3.81	88
Colour Word Interference Inhibition/Switching	6.29a	3.86	7	8.31a	3.25	74
Letter Fluency	9.33a	4.63	6	8.97a	3.26	76
Category Fluency	7.63a	4.81	8	10.24a	3.67	76
Category Switching	7.25a	3.99	8	9.00a	3.31	71
Tower Test	9.67a	2.42	6	10.61a	2.17	74

Note: SD = Standard Deviation. Significant differences are highlighted in bold italics

The final TBI sample included 90 participants, with 10 removed due to the failure on the PVTs. As the TBI sample was a clinical sample not all participants were administered all DKEFS subtests and, as such the TBI sample size can be seen to vary across the analyses.

Twenty-six orthopaedic patients were recruited from the Trauma and Orthopaedic outpatient clinic of the University Hospital Birmingham, UK. The participants were recruited on the basis that they had a history of non-head injury trauma within the last three years defined as long bone injuries or pelvic injuries with no history of previously sustained head injury or having experienced a loss of consciousness. Participants were screened for eligibility (Appendix III). Those with severe and enduring mental health problems (e.g. schizophrenia), serious neurological degenerative disorders (e.g. Alzheimer's dementia), learning disability or pervasive developmental conditions (e.g. autistic spectrum) which could cause impairment on cognitive tests, were excluded from the study. None of the orthopaedic controls failed the PVT. During testing, however, one orthopaedic control did not complete all the D-KEFS subtests administered.

Of the hundred people with TBI 21 (21%) were female and 79 (79%) were male, whereas for the orthopaedic controls 14 (53.8%) were female and 12 (46.2%) were male. The gender ratio was statistically significant between people with TBI and orthopaedic controls (Pearson Chi square = 11.097 p = 0.001) with males being over-represented in the TBI group. As the age of participants were presented in ranges, a parametric comparison of age was not calculated. When comparing the frequency of participants falling into each age category by group there was no significant difference in ages between people with TBI and orthopaedic controls (Pearson Chi square = 11.850 p = 0.106). Within the final TBI sample (after the removal of PVT Failures), 6 (6.7%) had an injury that was classified as mild-uncomplicated TBI; 16 (17.8%) suffered a mild-complicated TBI; 29 (32.2%) had a moderate TBI and 39 (43.3%) experienced a severe TBI. Regarding the orthopaedic group, 13 (50%) had broken ankle, leg, foot or toes, 7 (27%) had injuries to the arm or elbow, 4 (15%) were diagnosed with deformity of bone structures, and 2 (8%) were reported having musculoskeletal abnormalities. Premorbid intellectual status for TBI and orthopaedic patients was estimated from the Test of Premorbid Functioning UK (TOPF [UK]). There was a significant difference between the TBI and orthopaedic patients on TOPF [UK] predicted IQ ($t_{120} = -4.4$, p <0.01), with people with TBI having a lower 'premorbid' IQ status. The characteristics of people with TBI and orthopaedic controls are shown in Table 6.

Table 6: Characteristics of people with TBI and orthopaedic controls

	Traumatic Brain Injury		Orthopaedic	
	Frequency	%	Frequency	%
Age Group (Years)				
16-19	16	16	1	3.8
20-49	66	66	17	65.4
50-89	18	18	8	30.8
Gender		%		%
Male	79	79	12	46.2
Female	21	21	14	53.8
Education Level		%		%
9-11 years	51	51	6	23.1
12 years	11	11	0	0
13-15 years	27	27	5	19.2
16 or more years	11	11	15	57.7
Premorbid FSIQ Estimate	Mean	SD	Mean	SD
	96	9.1	105.48	11.650
	(N= 90)		(N = 26)	
Days since injury	295.5	199.3		
	(N= 42)			
Duration of PTA (in days)	7.44	8.74		
	(N= 70)			
Overall severity of injury	Frequency	%		
Mild Uncomplicated	6	6.7		
Mild/ Complicated	16	17.8		
Moderate	29	32.2		
Severe	39	43.3		
	(N= 90)			
Initial GCS				
13-15	26	28.9		
9-12	6	6.7		
3-8	28	31.1		

Note: GCS = Glasgow Coma Scale; PTA = Post Traumatic Amnesia

Measures

Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan & Kramer, 2001)

In this study the participants were administered the Trail Making Test, Colour-Word Interference

Test (Inhibition and Inhibition/ Switching), Verbal Fluency Test (Letter Fluency, Category

Fluency and Category Switching) and the Tower Test according to standard manual instructions

(Delis, Kaplan & Kramer, 2001b). Each of these has its own derived scores, with these scores

representing the indicators of executive functioning ability specific to each subtest. The

description of the DKEFS subtests administered in the study is listed in Table 7, along with the

related reliability and the derived scores utilised.

Table 7: D-KEFS subtests used in this study

DKEFS Subtest	Description	Domains measured	Derived Scores	Reliability
Trail Making	The primary executive function task is Number-Letter Switching in which the respondent is asked to alternate joining numbers and letters in order on an A3 array under timed conditions.	Cognitive Flexibility and Attentional Switching (Divided Attention) Baseline tasks (Conditions 1-3) control for visual selective attention, visual scanning, speed, and sequencing whilst condition 5 controls for simple motor speed	Completion-time score: Speed of completion Errors score: Total errors including sequencing or set loss errors	Adequate 0.6–0.81 (combined Number and Letter Sequencing Composite Score)
Colour Word Interference Test condition 3 (Inhibition)	The respondent is required to name the ink colour of words at speed when the word (e.g. RED) is dissonant with the ink colour in which the word is printed (e.g. GREEN).	Inhibition of an automatic (pre-potent) response. Baseline tasks of basic colour name at speed and word reading at speed control for more fundamental cognitive impairments.	Completion-time score: Speed of completion Error score: Total errors including self-corrected and uncorrected errors	Adequate 0.7–0.79
Colour Word Interference Test condition 4 (Inhibition/ Switching)	The respondent is required to inhibit as per condition 3 unless the word appears in a box in which case the respondent must read the word	Inhibition Switching measuring inhibition and cognitive flexibility.	Completion-time score: Speed of completion Error score: Total errors including self-corrected and uncorrected errors	Marginal 0.6–0.69
Verbal Fluency Test condition 1 (Letter Fluency)	The respondent is asked to generate as many words as possible that begin with a designated letter (F, A and S).	Letter Fluency (Phonemic Fluency) Generation of a verbal response and retrieval of information from semantic memory as well as perseveration and self-monitoring.	Total correct score: Number of correct words generated Error score: Set loss errors, repetition errors	High 0.68–0.90
Verbal Fluency Test condition 2 (Category Fluency)	The respondent is asked to generate as many words as possible, belonging to a designated semantic category but can begin with any letter designated letter across 2 trials.	Category (Semantic) Fluency Generation of a verbal response and retrieval of information from semantic memory as well as perseveration and self-monitoring. Acts as a baseline for the other verbal fluency tasks	Total correct score: Number of correct words generated Error score: Set loss errors, repetition errors	Marginal to adequate 0.53–0.76
Verbal Fluency Test condition 3 (Category Switching)	The respondent is asked to generate as many words as possible belonging to two semantic categories as well as alternating between categories in a 60 second time limit.	Category Switching Cognitive flexibility and self-monitoring	Total correct score: Number of correct words generated Total switching accuracy: Low Scores reflecting a failure to switch Error score: Set-loss errors, repetition errors	Total Correct: Low to marginal 0.37–0.68 Total switching: Marginal to adequate 0.51–0.76
Tower Test	The respondent is required to move disks across 3 pegs to make a specified pattern (Tower) using as few moves as possible whilst not breaking the task rules	Spatial planning, rule learning, inhibition of impulsive or perseverative responding and maintaining the instructional set	Total Achievement Score: Number of towers completed and the number of moves used Secondary measures include move accuracy ratio, time per move and rule violations.	Total achievement: Internal consistency ranged from 0.43–0.84 based on age group

Test of Premorbid Functioning UK (TOPF [UK]; Wechsler, 2011)

The TOPF[UK] is a revised and updated version of the Wechsler Test of Adult Reading which serves as an effective method for estimating pre-injury intellectual functioning (full-scale IQ) for individuals aged 16-89 before injury. This test is co-normed with the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) and Wechsler Memory Scale-Fourth Edition (WMS-IV). Therefore, the TOPF[UK] raw scores could be, as per manualised instructions, combined with demographic variables (age, gender and years of education) to generate predicted intellectual scores. Since executive performance is associated with intelligence, the effects of general intellectual level derived from this test are considered to increase the specificity of the findings and evaluate their impact on D-KEFS performance. On this reading test, the participants were required to read aloud a list of 70 words with irregular grapheme-phoneme translation. The TOPF was administered to both TBI and orthopaedic control participants.

Performance Validity Test (PVT)

PVTs are reported to be strongly related to the performance on neuropsychological tests (Lange, Pancholi, Bhagwat, Anderson-Barnes & French, 2012; Greve & Bianchini, 2007; Meyers, Volbrecht, Axelrod & Reinsch-Boothby, 2011; Green, Rohling, Leed-Haley & Allen, 2001). Of significance, the inclusion of the test is also recommended by the National Academy of Neuropsychology in the United States and the British Psychological Society Division of Neuropsychology. It is now a requirement for American Psychological Association (APA) journals that a test of performance validity be administered when a battery of neuropsychological tests is used. For this reason, the current study included PVTs as a measure to identify participants who are not applying appropriate effort on the neuropsychological tests

and this will be used in the analysis of the results. The Green's Word Memory Test (WMT; Green, Allen & Astner, 1995) and Test of Memory Malingering (TOMM; Tombaugh, 1997) were utilised in current study, using cut-offs provided in the test manuals. Both are well-validated PVTs as evidenced by the high specificity in TBI groups and are the most widely used PVTs among neuropsychologists (LaDuke, Barr, Brodale & Rabin, 2018). Failure as defined by the original test cut-offs specified in the PVT manuals, on a single stand-alone PVT resulted in the participant's data being judged invalid, were excluded from further analysis. Both WMT and TOMM were administered to the TBI group. Only WMT was administered to the orthopaedic group due to time constraints.

Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV)

The WAIS-IV is a highly reliable and commonly used measure of general intelligence (full-scale IQ). It also provides an estimate of general intellect ability (General Ability Index, GAI) with a reduced emphasis on the elements of working memory and processing speed than the full-scale IQ, given these constructs are more vulnerable to people with acquired brain injury. The WAIS-IV also generates other indices including Verbal Comprehension Index, VCI (verbal knowledge and reasoning); Perceptual Reasoning Index, PRI (non-verbal/visuospatial problem solving); Processing Speed Index, PSI (psychomotor speed) and Working Memory Index, WMI (working memory). The four subtests from the WAIS-IV administered to the participants with TBI were as follows: Vocabulary and Information (for the estimation of VCI); Matrix Reasoning and Block Design (for the estimation of PRI) from which a GAI was constructed. Both people with TBI and orthopaedic controls completed the Processing Speed subtests of the WAIS-IV (Coding and Symbol Search). They also completed the Letter Number Sequencing/

Digit Span task from the WAIS-IV. However, orthopaedic controls were not administered tests which assessed their current intellectual functioning due to time constraints.

Procedures

People with TBI were administered the TOPF, Word Memory Test, TOMM, DKEFS subtests (Colour Word Interference Test, Verbal Fluency Test, Trail Making Test and Tower test) as well as four WAIS-IV subtests. Tests were administered to patients who had sustained a Traumatic Brain Injury (TBI) and had been referred to the outpatient TBI clinic at a UK major trauma centre. It is noted that data for the TBI sample was gathered as part of routine clinical assessment by clinical staff at University Hospital Birmingham. The severity of TBI was classified by a clinical categorisation using initial Glasgow Coma Scale (GCS) score (post-resuscitation if appropriate), duration of Post Traumatic Amnesia (PTA), duration of loss of consciousness and imaging results if available. Full details are provided in the Appendix IV. Patients were excluded if they were still experiencing PTA before the time of testing, as judged by Consultant Clinical Neuropsychologist during clinical interview and in-patient records where applicable. Psychometric testing was carried out either by a Consultant Clinical Neuropsychologist or an Assistant Psychologist under supervision.

Orthopaedic control participants were recruited from the outpatient Trauma and Orthopaedic clinic at University Hospital Birmingham by a Trainee Clinical Psychologist. Prior to recruitment, the invitation letter as well as the information sheet which gave detailed information about the study were delivered to all orthopaedic patients by the Trainee. Letters were sent with the help of the Trauma and Orthopaedic outpatient secretary team at least two days before the patients' orthopaedic appointments. Therefore, all potential participants should

have a minimum of 48 hours to decide if they wished to take part in the study. On the day of the clinic appointment, the Trainee approached and identified potential participants in the orthopaedic clinic waiting area based on the eligibility criteria established. If the potential participant expressed his/ her interests in participating the study, further explanation was given to ensure each participant understood the study.

Participants were only tested when they understood the study and had given informed consent for their participation (see Appendix V for details). Tests were administered in the following order: TOPF; DKEFS Verbal Fluency (Letter Fluency, Category Fluency and Category Switching); Word Memory Test; WAIS-IV Processing Speed (Coding and Symbol Search); DKEFS Trail Making (Number Sequencing, Letter Sequencing, Number-Letter Sequencing); Colour Word Interference (Colour Naming, Inhibition and Inhibition/ Switching); WAIS-IV Letter-Number Sequencing and Tower Test. Each testing session took between 90 to 120 minutes to complete and was administered at the neuropsychology outpatient clinic at University Hospital Birmingham by the Trainee Clinical Psychologist.

Both the orthopaedic control group and the TBI group, were administered additional measures of performance validity, premorbid intellectual functioning and mental processing speed to account for other influences on D-KEFS performance that were not specific to executive functioning. The base rates of low scores in people with TBI relative to orthopaedic controls were also examined at a range of percentile cut-offs since it is likely that a TBI sample will be heterogeneous with regard to the degree of cognitive impairment. This was intended to provide important data with which to identify and quantify true deficits in executive functioning from normal performance variability.

Ethical approval

The ethical approval of this study was granted by the West Midlands-Edgbaston Research

Ethics Committee (REC), the Health Research Authority (HRA), as well as by the R&D

department at University Hospitals Birmingham (see Appendix VI).

RESULTS

Mean differences in the TBI and orthopaedic groups on different D-KEFS measures

Means and standard deviations of the TBI and orthopaedic control groups on different D-KEFS

executive measures are presented in Table 8. Statistical probabilities were corrected for

multiple comparisons using the Bonferroni correction. The orthopaedic controls scored

significantly higher than the TBI group on the Trail Making Number-Letter Switching, Colour-

Word Interference Inhibition and Inhibition/ Switching, Letter Fluency and Category

Switching. These comparisons also revealed significant group differences with large effect

sizes ($p < 0.05$, $0.73 <$ Cohen's $d < 1.08$), indicating significant impairments of executive

functions in people with TBI relative to the orthopaedic controls. The Colour Word Inhibition/

Switching evidenced the greatest difference in performance of more than one standard deviation

whereas the Letter Fluency showed the least difference. Both the Category Fluency and Tower

Test, however, failed to reach statistical significance.

Table 8: Univariate differences between the TBI and Orthopaedic groups on the D-KEFS measures of executive functioning

D-KEFS subtests	Traumatic Brain Injury			Orthopaedic			
	N	Mean	SD	N	Mean	SD	Cohen's d
Trail Making Number-Letter Switching	89	8.43$_a$	3.65	26	11.31$_b$	2.11	0.97
Colour Word Interference Inhibition	88	9.07$_a$	3.81	25	11.64$_b$	1.98	0.85
Colour Word Interference Inhibition/Switching	74	8.31$_a$	3.25	25	11.32$_b$	2.25	1.08
Letter Fluency	76	8.97$_a$	3.26	25	11.24$_b$	2.95	0.73
Category Fluency	76	10.24$_a$	3.67	25	11.72$_a$	3.37	0.42
Category Switching	71	9.00$_a$	3.31	25	11.96$_b$	3.03	0.93
Tower Test	74	10.61$_a$	2.17	25	10.88$_a$	1.90	0.13

Note: SD = Standard Deviation

Values in the same row not sharing the same subscript are significantly different at p< .05 in the two-sided test of equality for column means. Cells with no subscript are not included in the test. Tests assume equal variances.[1]

[1] Tests are adjusted for all pairwise comparisons with a row of each innermost subtable using Bonferroni correction.

Construction of the Executive Functioning Indices (EFIs)

The D-KEFS Executive Function Index (EFI) was calculated using the procedures described in Crawford, Garthwaite, Sutherland & Borland (2011). Two EFIs were computed using different combinations of the D-KEFS subtests. The EFI-A was constructed from all available D-KEFS subtests administered whereas the EFI-B excluded the Tower Test due to its reduced sensitivity to TBI. The EFI-A and EFI-B were calculated using the formula as shown below:

$$\text{Executive Functioning Index} = 15/SD_{Sum} \cdot (X_{Sum} - \bar{X}) + 100$$

Where, X_{Sum} is the sum of the achievement scores for an individual participant; SD_{Sum} is the population standard deviation and $\bar{X}$ is the population mean of the sum of the achievement scores respectively.

For instance, in the calculation of EFI-A, X_{Sum} is the sum of each participant's achievement scores of the seven available D-KEFS subtests (Trail Making Number-Letter Switching; Colour Word Interference Inhibition; Colour Word Interference Inhibition/ Switching; Letter Fluency; Category Fluency; Category Switching; and Tower Test). SD_{Sum} was estimated from the performance of the D-KEFS normative data and the population mean $\bar{X}$ is 70 given that each individual's achievement scores has a mean of 10. After applying the procedure the composite EFI score has a mean of 100 and a standard deviation of 15.

Differences in performance between people with TBI and orthopaedic controls on the EFIs

Table 9 shows the comparison of performance between the TBI and orthopaedic control group on EFIs. The two groups significantly differed, indicating that an advantage existed for orthopaedic controls over people with TBI ($p<0.05$) on both the EFI-A and EFI-B with large effect sizes in both. As can be seen in the table below, the effect sizes were greater for composite executive indices than for individual components.

Table 9: Univariate differences between the TBI and Orthopaedic groups on EFIs

	Traumatic Brain Injury			Orthopaedic			
	N	Mean	SD	N	Mean	SD	Cohen's d
EFI-A	64	94.87$_a$	17.21	25	111.69$_b$	11.30	1.16
EFI-B	66	93.92$_a$	17.75	25	111.63$_b$	11.65	1.94

Note: SD = Standard Deviation.

Values in the same row not sharing the same subscript are significantly different at p< .05 in the two-sided test of equality for column means. Cells with no subscript are not included in the test. Tests assume equal variances.[1]

D-KEFS performances controlling for premorbid intellectual status and processing speed

To increase the specificity of findings to executive functioning, analysis of covariance (ANCOVA) was calculated with premorbid intellectual status as a covariate. A significant

difference between participant groups was observed for EFI-A (F = 5.97, p = 0.017) with a medium effect size (Partial Eta squared = 0.065) which is equivalent to a Cohen's d = 0.53. For EFI-B the difference remained significant (F = 6.62, p <0.012), with a similarly medium effect size observed (Partial Eta squared = 0.070; Cohen's d equivalent = 0.55). When processing speed was controlled for, all significant differences on the EFIs vanished. A non-significant difference between participant groups was observed for both EFI-A (F = 1.81, p = 0.18) and EFI-B (F = 2.02, p = 0.16).

The correlations between Processing Speed Index (PSI) and measures of executive functioning are shown in Table 10. Significant correlations were found between processing speed and measures of executive functioning in the orthopaedic control group on Colour Word Interference Inhibition, Category Fluency, Category Switching, Trail Making Number-Letter Switching and the two Executive Functioning Indices (EFI-A & -B).

Table 10: Correlations between PSI and measures of executive functioning

		Colour Word Interference Inhibition	Colour Word Inhibition/ Switching	Tower Test	Letter Fluency	Category Fluency	Category Switching	Trail Making Number-Letter Switching	EFI-A	EFI-B
TBI PSI	Pearson Correlation	.142	.076	.011	.073	.148	.151	.136	.143	.054
	Sig. (2-tailed)	.285	.603	.938	.608	.316	.244	.372	.348	.701
	N	59	49	53	52	48	61	45	45	52
Orthopaedic PSI	Pearson Correlation	.432*	.283	.008	.228	.535**	.459*	.651**	.663**	.463*
	Sig. (2-tailed)	.035	.180	.972	.283	.007	.021	.001	.000	.023
	N	24	24	24	24	24	25	24	24	24

Note: Significant correlations are highlighted in bold.

Discrimination of people with TBI from orthopaedic controls based on EFIs

The area under the receiver operating characteristic curve (AUC) described in Table 11

quantifies the performance of the EFIs in discriminating people with TBI from orthopaedic

controls. As can be seen in the table below, both EFI-A and EFI-B reliably discriminated

between the TBI and orthopaedic groups. Both indices showed significantly superior

classifications, with the EFI-A demonstrating an AUC of 0.813 which is equivalent to a

Cohen's d = 1.26 (i.e. a large effect size) whereas the EFI-B demonstrated an AUC of 0.819

which is equivalent to a Cohen's d around 1.30 (i.e. a large effect size) (Rice & Harris, 2005).

Table 11: Area under the curve, standard error and 95% confidence intervals (CIs) for the discrimination between the TBI and orthopaedic groups

Executive Functioning Indices	AUC	Standard Error	Asymptotic Sig.	Asymptotic 95% CI	
				Lower Bound	Upper Bound
EFI-A	0.813	.048	0.000	0.719	0.906
EFI-B	0.819	.046	0.000	0.728	0.910

Note: AUC = Area Under the ROC Curve

EFIs performance by TBI severity

The performance of participants with mild uncomplicated/ complicated TBI on EFI-A and EFI-

B was compared with performance of participants with moderate-to-severe TBI to examine the

effect of injury severity on Executive Functioning Indices. Table 12 shows the comparison of

performance between the mild TBI and moderate-to-severe TBI groups on the two EFIs. Only

the EFI-A evidenced significant differences in mean performance, with the mild TBI group

significantly scoring higher than the moderate-to-severe TBI group on this index. The group

difference was also quantified by a medium effect size of Cohen's d = 0.77. It therefore

appears that the EFI-A is more sensitive to the severity of TBI than the EFI-B, showing that the

EFI-A demonstrated a reasonable relationship to TBI severity (i.e. severity gradient).

Table 12: EFI mean differences for mild versus moderate-severe TBI

	Mild Uncomplicated-complicated TBI			Moderate-Severe TBI			
	N	Mean	SD	N	Mean	SD	Cohen's d
EFI-A	11	103.89a	13.03	57	91.8b	17.97	0.77
EFI-B	14	100.93a	14.34	57	90.6a	18.55	0.62

Note: SD = Standard Deviation

A Receiver Operating Characteristic (ROC) analysis was also calculated to quantify the discriminative ability of the EFIs in differentiating between TBI severity groups. The results are illustrated in Table 13. Both the EFI-A and EFI-B performed less well when they were used to discriminate between individuals in the mild TBI group and those in the moderate to severe TBI group. The EFI-A demonstrated an AUC of 0.764 which is equivalent to a Cohen's d = around 1, i.e. a large effect size whereas the EFI-B demonstrated an AUC of 0.751 which is equivalent to a Cohen's d around 0.96, i.e. a large effect size (Rice & Harris, 2005). The area under the curve values for both EFIs are considered to be 'fair' at correctly classifying those with mild severity from those with moderate-to-severe TBI, suggesting that the EFIs may be comparatively less sensitive to the severity of TBI.

Table 13: Area under the curve for the discrimination between mild uncomplicated-complicated and moderate-severe TBI

Executive Functioning Indices	AUC	Standard Error	Asymptotic Sig.	Asymptotic 95% CI	
				Lower Bound	Upper Bound
EFI-A	0.764	.074	0.012	0.618	0.909
EFI-B	0.751	.077	0.017	0.599	0.902

Note: AUC = Area Under Curve

Comparing the base rates of D-KEFS performance in orthopaedic controls to people with TBI

Crawford, Garthwaite, Sutherland & Borland (2011) and Karr, Garcia-Barrera, Holdnack & Iverson (2018) have previously examined the base rates of low scores at different percentile

cut-offs to establish the degree of abnormality across the DKEFS subtests using normative data, in the latter case further stratifying the data by years of education and level of intelligence. A similar approach was adopted in this study to compare the prevalence of low D-KEFS performance in orthopaedic controls to people with TBI.

Table 14 provides the percentage of people with TBI and orthopaedic controls on the D-KEFS performance at or below selected percentile cut-offs. When examining the multivariate base rates of low scores across all of the DKEFs subtests that were administered, the TBI group presented with a consistently higher rate of obtaining low scores in comparison to the orthopaedic controls. As can be seen in Table 14, people with TBI were about twice as likely to have four D-KEFS subtests scored at $\leq 25^{th}$ percentile; or one subtest at $\leq 16^{th}$ percentile when all the D-KEFS subtests administered were considered simultaneously. It is also worth noting that 52% of the orthopaedic controls performed poorly on one D-KEFS subtest at low scores $\leq 25^{th}$ percentile. This indicates that it is not uncommon for people with orthopaedic injury to obtain some low scores when completing neuropsychological tests. In addition, the higher prevalence of attaining low scores also appears to be more evident when the administration is limited to a single subtest.

Table 14: The performance of the orthopaedic and TBI groups on the D-KEFS subtests at or below selected percentile cut-offs

Number of subtests failed at or below the stated percentile							
≤ 25th percentile	1	2	3	4	5	6	7
Orthopaedic Patients	52.00%	24.00%	8.00%	8.00%	0.00%	0.00%	0.00%
TBI	63.62%	38.62%	27.69%	15.19%	10.94%	1.56%	0.00%
Odds Ratio	1.22	1.61	3.46	1.90	NA	NA	NA
< 16th percentile	1	2	3	4	5	6	7
Orthopaedic Patients	24.00%	4.00%	0.00%	0.00%	0.00%	0.00%	0.00%
TBI	51.02%	29.14%	19.77%	5.70%	3.13%	1.56%	0.00%
Odds Ratio	2.13	7.29	NA	NA	NA	NA	NA
≤ 9th percentile	1	2	3	4	5	6	7
Orthopaedic Patients	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
TBI	37.30%	24.80%	13.87%	1.37%	0.00%	0.00%	0.00%
Odds Ratio	NA	NA	NA	NA	NA	NA	NA
≤ 5th percentile	1	2	3	4	5	6	7
Orthopaedic Patients	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
TBI	33.12%	19.05%	4.99%	0.30%	0.00%	0.00%	0.00%
Odds Ratio	NA	NA	NA	NA	NA	NA	NA

The base rates of low scores across both of the EFIs were then examined. Similarly, there was a consistently higher rate of low scores in the TBI group when compared to the orthopaedic controls on EFIs performance. Table 15 shows the percentage of orthopaedic and TBI groups scored on the two EFIs at different percentile cut-offs. When considering performance below 50th percentile, the TBI group was found to be 4 times more likely to obtain score at this level compared to the orthopaedic control group for both EFI-A and EFI-B. As none of the orthopaedic controls scored below 25th percentile or at lower percentiles on the two EFIs, it is therefore impossible to estimate the odds ratio of how likely people with TBI have low EFI scores at these levels than their orthopaedic counterparts.

Table 15: The performance of the orthopaedic and TBI groups on the EFIs at or below selected percentile cut-offs

		Percentile cut-offs					
		<50	<25	<16	<9	<5	<2
Standard Score equivalent		**<100**	**<90**	**<85**	**<80**	**<75**	**<69**
EFI-A							
Orthopaedic Patients	N=25	15.40%	0.00%	0.00%	0.00%	0.00%	0.00%
TBI	N=64	62.10%	34.80%	30.30%	25.80%	24.20%	12.10%
	Odds ratio	4.03	NA	NA	NA	NA	NA
EFI-B							
Orthopaedic	N=25	15.40%	0.00%	0.00%	0.00%	0.00%	0.00%
TBI	N=66	62.10%	42.40%	31.30%	23.40%	18.80%	9.40%
	Odds ratio	4.03	NA	NA	NA	NA	NA

DISCUSSION

In the present study a number of selected D-KEFS subtests were utilised to examine the validity of the D-KEFS in the evaluation of executive functioning in a group of patients with mild-uncomplicated to severe traumatic brain injury. In the first part of the study, the performance of the TBI sample on the selected D-KEFS subtests was compared with that of a group of acute orthopaedic controls with no head injury, a cohort which is considered as a better control group than a non-injury sample as it accounts for general trauma factors and injury-related experiences (e.g. hospitalisation and pain). The results demonstrated that TBI patients showed marked deficits in executive functioning compared to orthopaedic controls, with moderate to large effect sizes, ranging from 0.73 to 1.08. People with TBI performed significantly worse than the orthopaedic controls on the Trail Making Number-Letter Switching, Colour Word Interference Inhibition, Colour Word Interference Inhibition/ Switching, Letter Fluency and Category Switching tasks, indicating that the D-KEFS showed some criterion validity in discriminating executive impairment related to traumatic brain injury from normal performance. This finding is consistent with recent literature on TBI in support of the validity

of the D-KEFS (Anderson, Jaroh, Smith, Strong & Donders, 2017; Lengenfelder, Arjunan, Chiaravalloti, Smith & DeLuca, 2015; Heled, Hoofien, Margalit, Natovich & Agranov, 2012; Strong, Tiesma & Donders, 2011). Despite this, the Category Fluency and Tower Test did not yield any statistically significant group difference in this study. The findings may suggest that these subtests lack the sensitivity to the effect of TBI. Whilst there have been relatively few studies investigating the efficacy of Category Fluency and Tower Test specifically on the impact of TBI, some researchers have found that Category Fluency did not appear to be very sensitive in differentiating TBI patients from healthy controls (Anderson, Jaroh, Smith, Strong & Donders, 2017; Strong, Tiesma & Donders, 2011). Notably though, one study has reported the Tower Test demonstrating a significant group difference with large effect size in a group of TBI patients. However, only nine patients were recruited in that study (Ghawami, Sadeghi, Raghibi & Rahimi-Movaghar, 2017).

As discussed above, many of the individual D-KEFS subtests have been criticised as having low reliability values. To address this issue, the construction of Executive Function Index (EFI) was incorporated in this study following the procedures described in Crawford, Garthwaite, Sutherland & Borland (2011). The construction of this composite index confers the advantage of providing a more reliable measure than might be obtained using individual D-KEFS subtests. Two D-KEFS EFIs were calculated based on all subtests that were administered (i.e. EFI-A) as well as those to be more discriminative of TBI relative to orthopaedic controls (i.e. EFI-B). Significant group mean differences were found on both the EFI-A and EFI-B, with TBI patients performing worse than the orthopaedic controls. More importantly the composite EFIs resulted in greater effect sizes than individual subtests.

To increase the specificity of the findings to executive functioning, the domains of premorbid intellectual status and processing speed were also controlled for. This additional analysis was considered important in order to examine whether the performance differences between people with TBI and orthopaedic controls would be accounted for by overall intellectual functioning and more fundamental cognitive abilities such as processing speed or working memory. With the effect of premorbid intellectual status being removed, significant differences on the EFIs still remained. This suggests that differences between groups in the EFIs were not fully accounted for by prior intellectual functioning. However, all significant differences in EFI scores between the groups vanished when the processing speed was controlled for. Of additional interest, when correlations between processing speed and measures of executive functioning were further examined, significant correlations were found only in the orthopaedic control group on Colour Word Interference Inhibition, Category Fluency, Category Switching and EFIs. This loss of significance suggests that processing speed accounts for a significant amount of variance in EFI performances.

A possible explanation for the results could be that EFIs and PSI measure different cognitive components but that EFIs could be functionally dependent upon the integrity of processing speed. For instance, sufficient processing speed is required for real time monitoring and switching of attentional resources when engaging in D-KEFS tasks. As such, the differences in performance on the EFIs were negated when the effect of the processing speed was removed because the executive functioning domains may be secondary to and reliant upon attentional resources and processing in working memory. Furthermore, the EFIs demonstrated superior discriminations between people with TBI and orthopaedic controls ($p = 0.000$, AUC = 0.813 for EFI-A; AUC = 0.819 for EFI-B). However, the discriminative ability of both the EFI-A and

EFI-B reduced when used to classify those with mild severity from those with moderate-to-severe TBI, indicating that the EFI measures are comparatively less sensitive to the severity of TBI.

Considering people with TBI might have variable performances among tests due to the heterogeneous range of injury severity measured with Glasgow Coma Scale, duration of Post Traumatic Amnesia as well as recovery within the group, comparison of mean performance differences may not be able to sufficiently capture and quantify the deficits in executive functioning due to TBI. In this study the approach of comparing the base rate of low scores in orthopaedic controls to that of people with TBI was adopted. The results indicated that the TBI group presented with a consistently higher rate of obtaining low scores in comparison to the orthopaedic controls. Additionally, people with TBI were about two times more likely to have four D-KEFS subtests scored at $\leq 25^{th}$ percentile; or one subtest at $\leq 16^{th}$ percentile when all the D-KEFS subtests administered were considered simultaneously. Notably there was 52% of the orthopaedic controls performed poorly on one D-KEFS subtest at low scores $\leq 25^{th}$ percentile, indicating that healthy individuals may also obtain low scores when completing neuropsychological tests (Karr, Garcia-Barrera, Holdnack & Iverson, 2018). Moreover, the higher prevalence of attaining low scores appears to be more evident when the administration is limited to a single subtest. This highlights the importance of administering multiple tests when assessing executive functioning to allow for enhanced reliability (Suchy, 2016).

In addition, similar pattern was also observed on EFI scores. There was a consistently higher rate of low scores in the TBI group when compared to the orthopaedic controls on EFIs performance. The TBI group was four times more likely than the orthopaedic control group to

obtain EFI scores below 50th percentile. Notably none of the orthopaedic controls scored below 25th percentile or at lower percentiles on the two EFIs. The use of base rates therefore provides both a reliable and valid means of identifying and quantifying true deficits in executive functioning from normal variability in performance.

In the current study both the TBI and orthopaedic groups were screened using well-validated PVTs to exclude those who did not utilize appropriate effort during testing. This is particularly important in order to control for any confounding effects of performance validity. However, a number of limitations still need to be taken into account whilst considering the results of the study. Firstly, the gender variable was not matched between the two clinical groups, with the number of male participants significantly higher in the TBI group than in the orthopaedic group though this disproportion matches the characteristic of a TBI population in that there are normally more men than women. Secondly, the orthopaedic controls had significantly higher premorbid intellectual functioning than people with TBI and this might have contributed to some of the large effect sizes observed in the study. Furthermore, the sample size in this study, although larger than most TBI studies, is still relatively modest. Repetition of research study would therefore be of benefit to increase the generalisability of the findings. In addition, incorporating switching paradigms to several subtests is one of its features when developing the D-KEFS in order to increase the sensitivity of the test. In this instance the results showed that such switching tasks were among the subtests measuring executive functions that showed larger effect sizes (i.e. Colour Word Inhibition/ Switching: 1.08, Trail Making Number-Letter Switching: 0.97 and Category Switching: 0.93). However not all baseline tasks of the test could be administered due to time constraints. Since executive functioning depends on more fundamental cognitive processes (Suchy, 2016), further study may therefore need to consider

controlling for the effects of the fundamental/ non-executive conditions that are needed to perform their respective switching conditions of the test.

Clinical implications

The findings from current study indicate that the use of composite EFIs resulted in greater effect sizes than individual subtests. This highlights that such composite measure appears to be more valid and reliable than individual tasks. Clinicians are therefore encouraged to compute the composite score in their interpretations rather than solely relying on a single executive function test. In addition, the approach of comparing the base rate of low scores between people with TBI and orthopaedic controls was adopted in this study. This information is important to identify and quantify true deficits in executive functioning from normal variability in performance. It has been shown that it is not uncommon for individuals who are neurologically intact to obtain one or more low scores when completing a battery of neuropsychological tests. Considering how commonly some low scores can occur, it may be more useful if clinicians interpret the results of multiple scores at different cut-offs rather than using an absolute cut-off to define abnormality. Without such knowledge of the prevalence of low scores across multiple measures, it may become easy to over-interpret an isolated low test score. In fact, the variability in test performance and the presence of some low scores are not specific to any particular test battery or to a certain normative sample (Karr, Garcia-Barrera, Holdnack & Iverson, 2018). There are many other possible reasons of obtaining some low scores. Thus, it is important for clinicians to interpret low scores carefully in the context of multiple sources of client information such as clinical history, test results and neuroimaging data. Clinicians may consider using base rates in their interpretations to reduce the potential for misdiagnosis of cognitive impairment.

CONCLUSIONS

To the best of the author's knowledge, this is the first study in the UK that has compared the performance of a sample of patients with mild-uncomplicated to severe TBI, with that of orthopaedic controls using D-KEFS. Relative to the orthopaedic controls the TBI patients showed marked deficits in executive functioning and with moderate to large effect sizes. One of the Executive Functioning Indices also demonstrated a reasonable relationship in relation to TBI severity. The results of this study suggest that some D-KEFS subtests and the Executive Functioning Indices EFIs in particular are good measures for the identification of executive dysfunctions following TBI in the mild-uncomplicated to severe range. Future studies should consider a larger and more gender balanced orthopaedic population. It may also be useful to specifically examine the performances of people with TBI suffering head trauma of various aetiologies on D-KEFS, thereby capturing a more comprehensive cognitive profile for this population group. In conclusion, the findings support the use of the D-KEFS EFIs for the assessment of executive functioning in TBI population.

REFERENCES

Acquired brain injury 2016-2017 statistics based on UK admissions. (n.d.). *Headway Brain Injury Association Web site*. Retrieved March 23, 2019, from https://www.headway.org.uk/about-brain-injury/further-information/statistics/

Alvarez, J., & Emory, E. (2006). Executive Function and the Frontal Lobes: A meta-analytic review. *Neuropsychology Review*, 16(1), 17-42. doi:10.1007/s11065-006-9002-x

Anderson, L. B., Jaroh, R., Smith, H., Strong, C. A. H., & Donders, J. (2017). Criterion validity of the D-KEFS colour–word and verbal fluency switching paradigms following traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *39*(9), 890-899. doi: 10.1080/13803395.2016.1277513

Ardila, A. (2008). On the evolutionary origins of executive functions. *Brain and Cognition, 68*, 92–99. doi:10.1016/j.bandc.2008.03.003

Barkley, R. A. (2011). Is executive functioning deficient in ADHD? It depends on your definitions and your measures. *The ADHD Report*, *19*(4), 1-9. doi:10.1521/adhd.2011.19.4.1

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "Abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, *24*(1), 31-46. doi:10.1093/arclin/acn001

Calamia, M., Markon, K., & Tranel, D. (2013). The robust reliability of neuropsychological measures: Meta-analyses of test–retest correlations. *The Clinical Neuropsychologist*, *27*(7), 1077-1105. doi:10.1080/13854046.2013.809795

Chung, H. J., Weyandt, L. L, & Swentosky, A. (2014). The physiology of executive functioning. In S. Goldstein, & J. A. Naglieri (Eds.), *Handbook of executive functioning* (pp. 13-27). New York, NY: Springer

Corrigan, J. D., & Hammond, F. M. (2013). Traumatic brain injury as a chronic health condition. *Archives of Physical Medicine and Rehabilitation*, *94*(6), 1199–1201. doi: https://doi.org/10.1016/j.apmr.2013.01.023

Crawford, J. R., Garthwaite, P. H., Sutherland, D., & Borland, N. (2011). Some supplementary methods for the analysis of the Delis–Kaplan Executive Function System. *Psychological Assessment*, *23*(4), 888-898. doi:10.1037/a0023712

Crawford, J. R., Sutherland, D., & Garthwaite, P. H. (2008). On the reliability and standard errors of measurement of contrast measures from the D-KEFS. *Journal of the International Neuropsychological Society*, *14*(6), 1069-1073. doi:10.10170S1355617708081228

Delis, D. C., Kaplan, E., & Kramer, J. (2001a). *Delis-Kaplan Executive Function System: Technical manual.* San Antonio, TX: Psychological Corporation.

Delis, D. C., Kaplan, E., & Kramer, J. (2001b). *Delis-Kaplan Executive Function System: Examiner's manual.* San Antonio, TX: Psychological Corporation.

Delis, D.C., Kramer, J., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan Executive Function System: An update. *Journal of the International Neuropsychological Society*, *10*(2), 301-303. doi:10.1017/s1355617704102191

Department of Health Long-term Condition NSF Team (2005). *The National Service Framework for Long-Term Condition.* London: Department of Health. Retrieved from https://www.nai.ie/assets/80/018E05DD-E34C-329F-12D6CD5DCE26ECF3_document/dh_4105369.pdf

Fleminger, S., & Ponsford, J (2005). Long term outcome after traumatic brain injury. *British Medical Journal*, *331*, 1419-1420.

Fuster, J. M. (1993). Frontal lobes. *Current Opinion in Neurobiology*, *3*(2), 160–165.

doi:10.1016/0959-4388(93)90204-C

Ghawami, H., Sadeghi, S., Raghibi, M., & Rahimi-Movaghar, V. (2017). Executive
functioning of complicated-mild to moderate traumatic brain injury patients with frontal
contusions. *Applied Neuropsychology: Adult*, *24*(4), 299-307.

Goldstein, S., Naglieri, J. A., Princiotta, D., & Otero, T. M. (2014). Introduction: A history of
executive functioning as a theoretical and clinical construct. In S. Goldstein & J. A.
Naglieri (Eds.), *Handbook of executive functioning* (pp. 3-12). New York, NY:
Springer.

Green, P., Allen, L. M., & Astner, K. (1995). *The Word Memory Test: A manual for the oral
and computerized forms*. Edmonton, Canada: Green's Publishing Inc.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has a greater
effect on test scores than severe brain injury in compensation claimants. *Brain
Injury*, *15*(12), 1045-1060. doi:10.1080/02699050110088254

Greve, K. W., & Bianchini, K. J. (2007). Detection of cognitive malingering with tests of
executive function. In G. J. Larrabee (Ed.), *Assessment of malingered
neuropsychological deficits* (pp. 171–225). New York, NY: Oxford University Press

Heled, E., Hoofien, D., Margalit, D., Natovich, R., & Agranov, E. (2012). The Delis–Kaplan
Executive Function System Sorting Test as an evaluative tool for executive functions
after severe traumatic brain injury: A comparative study. *Journal of Clinical and
Experimental Neuropsychology*, *34*(2), 151-159. doi:10.1080/13803395.2011.625351

Homack, S., Lee, D., & Riccio, C. (2005). Test Review: Delis-Kaplan Executive Function
System. *Journal of Clinical and Experimental Neuropsychology*, *27*(5), 599-609. doi:
10.1080/13803390490918444

Iverson, G. L., Holdnack, J. A., & Lange, R. T. (2013). Using the WAIS-IV/WMS-IV/ACS

following moderate-severe traumatic brain injury. In J. A. Holdnack, L. W. Drozdick, L. G. Weiss, & G. L. Iverson (Eds.), *WAIS-IV, WMS-IV, and ACS: Advanced Clinical Interpretation* (pp. 485-544). New York, NY: Academic Press.

Jurado, M., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review*, 17(3), 213-233. doi: 10.1007/s11065-007-9040-z

Karr, J. E., Garcia-Barrera, M. A., Holdnack, J. A., & Iverson, G. L. (2018). Advanced clinical interpretation of the Delis-Kaplan Executive Function System: Multivariate base rates of low scores. *The Clinical Neuropsychologist*, *32*(1), 42-53. doi:10.1080/13854046.2017.1334828

Keifer, E., & Tranel, D. (2013). A neuropsychological investigation of the Delis-Kaplan executive function system. *Journal of Clinical and Experimental Neuropsychology*, *35*(10), 1048-1059.

Keil, K., Baldo, J., Kaplan, E., Kramer, J., & Delis, D. C. (2005). Role of frontal cortex in inferential reasoning: Evidence from the word context test. *Journal of the International Neuropsychological Society*, *11*(4), 426-433.

LaDuke, C., Barr, W., Brodale, D. L., & Rabin, L. A. (2018). Toward generally accepted forensic assessment practices among clinical neuropsychologists: A survey of professional practice and common test use. *The Clinical Neuropsychologist*, *32*(1), 145-164. doi:10.1080/13854046.2017.1346711

Lange, R. T., Pancholi, S., Bhagwat, A., Anderson-Barnes, V., & French, L. M. (2012). Influence of poor effort on neuropsychological test performance in U.S. military personnel following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 34*(5), 453–466. doi:10.1080/13803395.2011.648175

Lengenfelder, J., Arjunan, A., Chiaravalloti, N., Smith, A., & DeLuca, J. (2015). Assessing

frontal behavioral syndromes and cognitive functions in traumatic brain injury. *Applied Neuropsychology: Adult*, *22*(1), 7-15. doi:10.1080/23279095.2013.816703

Luria, A. R. (1966). *Higher cortical functions in man*. New York: Basic Books.

Luria, A. R. (1973). *The working brain*. London: Penguin.

McKinlay, W. W., & Brooks, D. N. (1984). Methodological problems in assessing psychosocial recovery following severe head injury. *Journal of Clinical and Experimental Neuropsychology*, *6*(1), 87-99. doi:10.1080/01688638408401199

Meyers, J. E., Volbrecht, M., Axelrod, B. N., & Reinsch-Boothby, L. (2011). Embedded symptom validity tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology*, *26*(1), 8-15. doi:10.1093/arclin/acq083

Miyake, A., Emerson, M., & Friedman, N. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. Seminars in Speech and Language, Volume 21(Number 02), 0169-0183. doi:10.1055/s-2000-7563

National Clinical Guideline Centre. (2014). *Head injury: triage, assessment, investigation and early management of head injury in children, young people and adults.* London: National Institute for Health and Care Excellence (UK). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK248061/pdf/Bookshelf_NBK248061.pdf

Puente, A. N., Lindbergh, C. A., & Miller, L. S. (2015). The relationship between cognitive reserve and functional ability is mediated by executive functioning in older adults. *The Clinical Neuropsychologist*, *29*(1), 67-81. doi:10.1080/13854046.2015.1005676

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1–38). Hove: Psychology Press.

Rabinowitz, A. R., & Levin, H. S. (2014). Cognitive sequelae of traumatic brain injury. *The Psychiatric Clinics of North America*, *37*(1), 1-11. doi:10.1016/j.psc.2013.11.004

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and Human Behaviour*, *29*(5), 615-620. doi:10.1007/s10979-005-6832-7

Schmidt, M. (2003). Hit or miss? Insight into executive functions. *Journal of the International Neuropsychological Society*, *9*(6), 962–964. doi:10.1017/s1355617703230162

Smith, C. (2011). Neuropathology. In J. M. Silver, T. W. McAllister, & S. C. Yudofsky (Eds.), *Text book of traumatic brain injury* (2nd ed., pp. 23-36). Arlington, VA: American Psychiatric Publishing, Inc.

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

Strong, C. A. H., Tiesma, D., & Donders, J. (2011). Criterion validity of the Delis-Kaplan Executive Function System (D-KEFS) fluency subtests after traumatic brain injury. *Journal of the International Neuropsychological Society*, *17*(2), 230-237.

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research*, *63*, 289–298. doi:10.1007/s004269900007

Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome? *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *362*(1481), 901–915.

Suchy, Y. (2016). *Executive functioning: A comprehensive guide for clinical practice*. New York, NY: Oxford University Press.

Swanson, J. (2005). The Delis-Kaplan executive function system: A review. *Canadian Journal of School Psychology*, *20*(1-2), 117-128. doi:10.1177/0829573506295469

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, *9*(3), 260-268. doi:10.1037/1040-3590.9.3.260

Tranel, D., Anderson, S. W., & Benton, A. (1994). Development of the concept of "executive function" and its relationship to the frontal lobes. In F. Boller, & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. *9*, pp. 125–148). Amsterdam: Elsevier.

Wechsler, D. (2011). *Test of premorbid functioning. UK version (TOPF UK)*. UK: Pearson Corporation.

World Health Organisation. (2006). *Neurological disorders: Public health challenges.* Switzerland: WHO Press. Retrieved from https://www.who.int/mental_health/neurology/neurological_disorders_report_web.pdf

Yates, P. J., Williams, W. H., Harris, A., Round, A., & Jenkins, R. (2006). An epidemiological study of head injuries in a UK population attending an emergency department. *Journal of Neurology, Neurosurgery & Psychiatry*, *77*(5), 699-701.

Yochim, B. P., Baldo, J. V., Kane, K. D., & Delis, D. C. (2009). D-KEFS Tower Test performance in patients with lateral prefrontal cortex lesions: The importance of error monitoring. *Journal of Clinical and Experimental Neuropsychology*, *31*(6), 658-663.

PUBLIC DOMAIN BRIEFING PAPER

Validation of the Delis-Kaplan Executive Function System (D-KEFS) in participants with Traumatic Brain Injury

by

Yin-Ming Chan

Department of Clinical Psychology

School of Psychology

The University of Birmingham

April 2019

THE CLINCIAL USEFULNESS OF DELIS-KAPLAN EXECUTIVE FUNCTION SYSTEM (D-KEFS) IN THE EVALUATION OF EXECUTIVE FUNCTIONS IN CLINICAL POPULATIONS WITH ACQUIRED NEUROLOGICAL CONDITONS

Introduction

Psychologists use the term 'executive functions' to describe the ability of planning, self-managing and organising. Issues with executive functioning can have a profound impact on many aspects of our everyday life, affecting us to live and work independently. It is recognised that executive functioning problems are common in people who have diseases or injuries caused to their brain systems such as head injury, brain tumour, stroke or dementia. Identifying problems in executive functioning is therefore important for psychologists to develop effective strategies to help these people improve their executive functioning skills. For this reason, a valid and reliable measure of executive function is essential. The Delis-Kaplan Executive Function System (D-KEFS) is a set of tests that comprehensively assess executive functions in both children and adults. Although there is some evidence in support of the effectiveness of the D-KEFS, its usefulness of identifying executive functioning problems in people with different brain diseases has not been systematically reviewed.

Aim

To systematically review all the published literature on the usefulness of the D-KEFS, specifically focusing on its ability to identify any executive functioning problems in people with different brain diseases or injuries from healthy individuals.

Method

A literature search in three databases was conducted. Search terms related to executive function, brain diseases or injuries, D-KEFS and clinical usefulness were collectively used to locate relevant journal articles. The resulting papers were reviewed, and rated for quality according to a set of quality criteria. Additional papers were found from the reference lists of these articles.

Results

The search strategy resulted in thirty-two journal articles being included in this review. There were five studies focused on examining the executive functions of individuals with traumatic brain injury; six examined patients with brain lesions; thirteen analysed on different types of dementia or other neurodegenerative disorders; and eight only focused on epilepsy. The selection of the D-KEFS subtests varied across the reviewed studies.

Conclusions

The D-KEFS appears to be a useful evaluation tool of executive functioning, based on the available evidence. The findings indicated that participants with various brain diseases or injuries performed significantly more poorly than healthy individuals on the test. The test performance was also associated with certain brain regions or, with the brain atrophy, which may arise as a consequence of the progression of the disease. Although the D-KEFS may have some value in identifying executive functioning problems in people with different brain diseases or injuries from healthy individuals, the evidence available is insufficient. More research may need to be conducted in future.

EMPIRICAL PAPER:

INVESTIGATING THE VALIDITY OF THE DELIS-KAPLAN EXECUTIVE
FUNCTION SYSTEM (D-KEFS) AS A NEUROPSYCHOLOGICAL ASSESSMENT
TOOL FOR EXECUTIVE FUNCTIONS IN THE TRAUMATIC BRAIN INHURY (TBI)
IN THE UK

Introduction

Executive functioning problems are commonly reported in people following Traumatic Brain
Injury (TBI). The Delis Kaplan Executive Function (D-KEFS) is a set of tests which measures
a wide range of abilities associated with executive functioning. Currently there is substantial
evidence to support the validity of D-KEFS as a useful instrument to identify deficits in
executive functioning. However there are only a few studies specifically studying TBI and
most have relatively modest sample sizes.

Aim

To investigate the validity of the D-KEFS by comparing the performance of a sample of
patients with TBI, to that of a group of people who have not sustained a head injury but have
had a non-head related trauma (e.g. broken leg), using selected D-KEFS subtests. This distinct
clinical population is considered to be a 'gold standard' control for studying TBI given that they
are arguable representative of the TBI group both demographically and psychosocially.

Method

One hundred patients with Traumatic Brain Injury (TBI) and twenty-six patients with non-head
related trauma were recruited from outpatient clinics at a major UK hospital. Participants were
assessed using a number of selected D-KEF subtests, as well as other widely utilised cognitive

tests which are used to estimate their intellectual functioning before injury and the effort they were putting on the tests.

Results

The participants with TBI performed significantly more poorly than the comparison group on several D-KEFS subtests, as reflected by the consistent high rate of obtaining low scores across subtests. Relative to the non-head related injury comparison group, people with TBI tended to experience more difficulty with inhibition and flexibility of thinking and these skills are known to be highly associated with executive functioning. Notably, the D-KEFS subtests did well in distinguishing people with TBI from the comparison group. Nevertheless, the measures preformed relatively less well in identifying those with mild TBI from those with moderate-to-severe TBI.

Conclusions

To our knowledge this is the first study in the UK that has compared the performance of a sample of patients with TBI, to that of a group of people with non-head related injury using D-KEFS. The findings support the use of the D-KEFS for the assessment of executive functioning in TBI population. Future study may consider using a larger and more gender-balanced comparison group and examine the performances of TBI individuals suffering head injury of various aetiologies on D-KEFS to capture a more comprehensive cognitive picture for this population group.

Volume I: Appendix I − Appraisal criteria for quality assessment of studies reviewed

Criteria	Developed based on von Elm et al (2007), Vandenbroucke et al (2007) and Coughian, Cronin & Ryan (2007)
Background/ rationale and study objectives	Has the scientific background been explained in the introduction and the rationale for the study been reported? Does the introduction give an overview of what is known/ not known about the topic and where the gaps in knowledge exist? Does it offer a balanced critical analysis of the literature? Have specific aims, objectives or pre-specified hypotheses been stated? If so, are they clear? Do they reflect the information in the literature review? Is the purpose of the study/ research problem clearly identified?
Study design	Are the key elements of the study design presented? Are all variables well defined? In comparison studies, is a control group included in the study?
Study setting and participants	Has the setting, locations, sources of recruitment been clearly described? Has the target population been clearly identified? Are the cases representative of a defined population? How were the sample selected? Are the eligibility criteria clearly identified? Are the participant characteristics clearly stated (e.g. demographic, diagnosis, co-morbid conditions, etc)? For case-control studies, is the rationale for the choice of cases and controls given? Has the matching criteria been stated? Were cases and controls matched appropriately?
Sample size	Does the study explain how the sample size was arrived at? Was there a sufficient number of cases/ controls selected? In comparison studies a group size of 15 or less will be considered weak; a group size of 15-25 will be considered sufficient; and over 25 will be considered strong.
Attrition	Are the numbers of individuals at each stage of the study reported (e.g. numbers potentially eligible, confirmed eligible, included in the study, etc)? Are reasons for non-participation/ withdrawal given at each stage?
Data collection	For each variable of interest, have sources of data and methods of measurement been described? Has the assessment measures adequately been described? Is the measure appropriate? How the measurement was conducted? Was information on the validity and reliability of testing instruments outlined? Were the measurement methods similar in the cases and controls?
Analysis/ results	What type of data and statistical analysis were undertaken? Were all statistical methods used clearly described, including those used to control for confounding? How many of the sample participated? How quantitative variables were handled for analysis? How missing data, matching of cases and controls were addressed? How strong was the association? Were the findings significant? Were potential confounding factors considered? Were the results adjusted for confounding?
Potential Bias	Were both direction and magnitude of any potential bias discussed? Were efforts to address these biases described?
Limitations	Were limitations of the study discussed, including any sources of potential bias or imprecision?
Interpretation and generalisability	Was an overall interpretation of the results cautiously provided, considering objectives, analyses and results from similar studies? Were the findings linked back to the literature review? Were hypotheses supported? Did the results of the study fit with other available evidence? Were recommendations for future research made? Were clinical implications of findings outlined? Was the generalisability (external validity) of the results discussed? Was the validity, generalisability and precision of results discussed in relation to other studies?

Volume I: Appendix II – Overview of the reviewed studies under the category of different acquired neurological conditions

Traumatic Brian Injury

Year	Authors, publication & source of participants recruitment	Measure(s) used in the study to assess executive functions	Study aims	Summary of main findings	Limitations
2017	Anderson, Jaroh, Smith, Strong & Donders Journal of Clinical and Experimental Neuropsychology People with TBI were recruited in a rehabilitation facility at an outpatient setting A matched group of healthy controls was obtained from the D-KEFS normative database	**D-KEFS:** 1. VFT (Letter Fluency; Category Fluency; Category Switching) 2. CWIT (Colour Naming; Word Reading; Inhibition; Inhibition/ Switching)	To investigate the clinical utility of the D-KEFS CWIT and VFT in assessing people with TBI To determine whether the findings could provide support for the results of a previous study (Strong, Tiesma & Donders, 2011) with regard to the clinical utility of the D-KEFS VFT	Performances on switching conditions are sensitive to some TBI-severity indicators such as length of coma and nature of lesions. Longer length of coma was associated with poor performance on the CWIT Inhibition/ Switching task, whereas the presence of diffuse lesions predicted worse performance in VFT Category Switching. People with moderate-to-severe TBI performed significantly worse than the mild-uncomplicated TBI group and healthy controls on both the D-KEFS CWIT Inhibition/ Switching and VFT Category Switching tasks. However, the mild-uncomplicated TBI and control groups did not differ significantly on these 2 tasks. Using a hierarchical linear regression analysis, the combined subtests, Inhibition/ Switching and Category Switching task, performed well in distinguishing the presence of cognitive impairment (WAIS-IV Processing Speed Index standard score ≤ 80 as pre-determined impairment cut-off) in participants with TBI, yielding a classification accuracy of 76.57% (sensitivity = 72%; specificity = 77%), with a likelihood ratio of 3.12 and an AUC of 0.83. However, the combined task did less well when it was used to identify people with moderate-severe TBI from those who sustained mild-uncomplicated injury and control participants. The model only yielded a modest classification accuracy of	The sample was predominantly Caucasian people. Different methods were used to determine the presence of diffuse lesions. Some important TBI-related severity were not included. Quality of education, current levels of anxiety and depression had not been taken into account.

				66.07% (sensitivity = 64%; specificity = 67%), with a likelihood ratio of 1.93 and an AUC of 0.70.	
2017	Ghawami, Sadeghi, Raghibi & Rahimi-Movaghar Applied Neuropsychology: Adult People with TBI were recruited from the neurosurgery ward of a hospital A matched group of normal controls was recruited from the community same as the participants with TBI	**D-KEFS:** 1. TMT (all 5 conditions) 2. VFT (Letter Fluency, Category Fluency, Category Switching) 3. DFT (Filled Dots, Empty Dots, Dot Switching) 4. ST (Free Sorting) 5. TQT 6. TT **BADS** (Rule Shift Cards, Key Search and Modified Six Element Task)	To investigate the executive functioning in a group of TBI patients using selected tests of the D-KEFS and BADS	Group comparisons revealed substantial impairments of executive functions in the frontal contusion patients. They performed significantly worse on all the executive measures (D-KEFS and BADS) when compared to normal controls, with very large effect sizes (p ≤ 0.003, 1.56 < Cohen's d < 3.12). When removing the effects of the fundamental conditions, the frontal group still performed significantly impaired on the switching conditions of the TMT, VFT and DFT. In terms of number of errors committed, the frontal patients committed considerably more total errors than the controls on the D-KEFS measures. Additionally, patients with lateral prefrontal contusions (LPFC) performed qualitatively worst on most of the measures, showing a significant tendency towards committing repetition/ perseverative errors. With regard to the lateralization effects, there were no significant group differences (as well as in committing errors) between patients with left- and right-frontal contusions on all executive measures with the exception of the VFT Letter Fluency where the left frontal group generated significantly less number of correct words than the right frontal group.	Small sample size. Absence of patients with non-frontal lesions as comparison group.
2016	Faber et al Brain Injury	**D-KEFS:** 1. VFT (Letter Fluency; Category Fluency; Category Switching)	To examine group differences in executive functioning between participants with and without early childhood	DTI analysis revealed the group with early moderate-to-severe TBI significantly exhibited decreased structural integrity of brain tissues in the left ventral striatum when compared to the control group of typically-developing children. On the measures of executive functioning, the	Modest sample size.

	No information about the sources from which the TBI and control groups were recruited	2. CWIT (Colour Naming; Word Reading; Inhibition; Inhibition/ Switching)	TBI on selected tests of the D-KEFS To investigate changes in neural structures after early TBI using two common Diffusion Tensor Imaging (DTI) metrics, Fractional Anisotropy (FA) and Apparent Diffusion Coefficient (ADC) as indicators of the white matter integrity, and to evaluate the effect of these changes in relation to executive functioning	TBI group showed significantly poorer performance on the Letter Fluency task in the VFT, the Inhibition and the Inhibition/ Switching tasks in the CWIT as well as more number of errors committed on the Inhibition task. Significant correlations were found between executive functioning and ventral striatum diffusion parameters. The right ($r = -0.524$, $p = 0.015$) and the left ($r = -0.566$, $p = 0.007$) ventral striatum ADC, and the left ($r = 0.474$, $p = 0.030$) ventral striatum FA were all associated with the performance on the VFT Category Switching task. The right ventral striatum FA was also correlated with the number of errors committed on Inhibition.	Great variability in the age at injury within the sample.
2012	Heled, Hoofien, Margalit, Natovich & Agranov Journal of Clinical and Experimental Neuropsychology People with TBI were recruited from a day treatment rehabilitation unit at a medical centre A matched control group of healthy participants was recruited from community groups	**D-KEFS ST** (Free Sorting, Sort Recognition) **TMT-A & -B** **WCST**	To investigate the ability of the D-KEFS ST, as compared with WCST and TMT, in differentiating between severe TBI patients and healthy control participants	The severe TBI group performed considerably worse than the control group across all 5 dimensions of the D-KEFS ST, namely "attempted sorts", "correct sorts", "free sorting description", "sort recognition description" and "perseveration sorting". Among these dimensions, "attempted sorts" was the most sensitive measure to distinguish group difference ($p \leq 0.001$, Cohen's $d = 1.11$) and also being one of the best predictors of group classification, with 9.5 sorts as the optimal cut-off point and an AUC of 0.8 (sensitivity = 88%; specificity = 65%). In contrast with D-KEFS ST, the WCST failed to identify group differences on all 6 measuring domains. The performance of the TBI group did not differ significantly from the healthy controls on aspects of the "perseveration responses" and "failure to maintain set" in the WCST. Nevertheless, TMT-B was the most	The two groups were not gender-balanced, with the control group having more females. Modest sample size.

				predictive test for classification of the TBI and the control groups.	
2011	Strong, Tiesma & Donders Journal of the International Neuropsychological Society People with TBI were recruited from a rehabilitation facility A matched group of healthy controls was obtained from the D-KEFS normative database	**D-KEFS:** 1. VFT (Letter Fluency, Category Fluency, Category Switching) 2. DFT (Filled Dots, Empty Dots, Dot Switching) **WCST**	To determine the criterion validity of the D-KEFS VFT and DFT in the neuro-psychological evaluation of patients with complicated mild-severe TBI	The TBI patients performed significantly worse than the control subjects only on the VFT Letter Fluency and Category Switching tasks, with very small effect sizes. Further logistic regression analysis indicated that the combined Letter Fluency and Category Switching tasks demonstrated a suboptimal classification accuracy of 65.39% (sensitivity = 66.15%; specificity = 64.62%), with a likelihood ratio of 1.87 and an AUC of 0.69. Findings also suggested that the D-KEFS DFT failed to discriminate between the TBI and control groups. None of the DFT tasks yielded statistically significant group differences. Longer length of coma significantly predicted poor performances on the Letter Fluency and Category Switching tasks. Further regression analysis indicated that the effect of injury severity on Letter Fluency performance was mediated in part by speed of information processing but this was not the case for Category Switching. Neither Letter Fluency nor Category Switching was correlated with WCST.	There were no patients with mild uncomplicated TBI in the sample, hence limiting the generalisability to this population group. The sample was predominantly Caucasian people.

Frontal Brain Lesions

Year	Authors, publication & source of participants recruitment	Measure(s) used to assess executive functions	Study aims	Summary of main findings	Limitations
2013	Keifer & Tranel Journal of Clinical and Experimental Neuropsychology The frontal groups were recruited from university hospitals and clinics No healthy controls	**D-KEFS:** 1. TMT (Number-letter Switching) 2. VFT (Letter Fluency, Category Fluency, Category Switching) 3. DFT (Filled Dots, Dot Switching) 4. CWIT (Inhibition, Inhibition/ Switching) 5. ST (Free Sorting, Sort Recognition) 6. TQT 7. WCT 8. TT 9. PT	To compare the performances of patients with vmPFC, dlPFC, and non-frontal lesions on the entire D-KEFS battery	On the 6 primary measures of the D-KEFS subtests (Category Fluency, Category Switching, Design Fluency Composite, Design Fluency Switching, Colour-Word Inhibition and Sort Recognition), patients with dorsolateral prefrontal cortex (dlPFC) brain lesions performed significantly impaired than patients with ventromedial prefrontal cortex (vmPFC) and non-frontal lesions (NF). However, the dlPFC group performed worse than the NF group but did not differ from the vmPFC group on measures of Colour-Word Inhibition/ Switching, Correct Sorts and Free Sorting Description. On almost every measure, mean scaled scores for the dlPFC group were below 9 whereas those in the vmPFC and NF lesion groups were 9 or above. However, there was no significant difference in performance between the vmPFC and the NF groups on any of the D-KEFS measures. There were no significant differences in group performance on any of the D-KEFS measures when the effects of FSIQ and processing speed were controlled for. The study finding also suggests that there were no significant group differences in the performances on measures with switching paradigms and their traditional counterparts.	Small sample size. Lack of racial diversity in the sample. Other indices such as measuring error rates could be considered.
2009	Yochim, Baldo, Kane & Delis	**D-KEFS TT**	To compare the performances of patients with focal lateral prefrontal cortex lesions	Patients with lateral prefrontal cortex (PFC) lesions performed significantly worse than controls on the D-KEFS TT. They completed fewer tower, spent longer time on each move, and committed more rule violations.	No comparison to a non-frontal control group.

	Journal of Clinical and Experimental Neuropsychology The frontal group was recruited from a medical facility No information about the source from which the control group was recruited		to control subjects on the D-KEFS TT	Additionally, participants who violated rules two or more times all had PFC lesions. Compared with only two control participants broke rules and each did so once, 10 out of the 12 frontal lesion participants committed two or more rule violations. Further analysis showed that the size of lesion correlated with the total rule violations committed ($r = 0.58$, $p = 0.05$). However, there was no significant group difference between PFC patients and controls with regard to the mean response time spent making the first move on each tower. Sensitivity and specificity were also calculated to assess the ability of the D-KEFS TT in the detection of frontal lesions. Committing two or more rule violations demonstrated 83% sensitivity and 100% specificity; a total achievement score of 14 or below resulted in 75% sensitivity and 83% specificity; if the mean amount of time that was spent on each move was 3.81 seconds or longer, this led to 75% sensitivity and 75% specificity; if participants completed seven or less towers, this brought about 75% sensitivity and 83% specificity.	Small sample size.
2007	Yochim, Baldo, Nelson & Delis Journal of the International Neuropsychological Society No information about the sources from which the frontal lesion and control groups were recruited	**D-KEFS TMT** (Visual Scanning, Number Sequencing, Letter Sequencing, Number-letter switching and Motor Speed)	To compare the performances of patients with focal lateral prefrontal cortex lesions to matched control subjects on the D-KEFS TMT	The frontal lesion group performed slower than the control group on the D-KEFS TMT, particularly on the Letter Sequencing and the Number-Letter Switching. When the performances of the 4 baseline conditions were controlled for, patients with frontal lesion were still disproportionately slower on the Number-Letter Switching condition. The frontal lesion group also showed a propensity to commit more errors in set-switching and sequencing on the switching condition than the control group. Compared with 4 out of the 11 controls, 9 out of the 12 patients with lateral PFC lesions made one or more errors.	Small sample size. No comparison to a non-frontal control group.

2005	Keil, Baldo, Kaplan, Kramer & Delis Journal of the International Neuropsychological Society No information about the sources from which the frontal lesion and control groups were recruited	**D-KEFS WCT**	To compare the performances of patients with prefrontal cortex lesions to matched control participants on the D-KEFS WCT	Patients with frontal lesions performed significantly impaired than control participants on the D-KEFS WCT. With regard to accuracy, patients with frontal lesions were less able to generate correct responses across the 10 trials. They also had to make more guesses to arrive at correct responses when compared to controls. Qualitatively, it was apparent that the frontal patients' responses were conceptually less appropriate within the context of a given sentence than controls. Frontal patients were also less able to integrate information across sentences on a given trial to infer the correct response. With regard to the lateralization effects, both the left- and right-frontal patients performed worse than controls. Effect size analysis revealed that patients with right-frontal lesions did better than those with left-frontal lesions on performances in terms of the accuracy, number of guesses and ability to integrate information.	Limited generalisability due to strict inclusion criteria. Small sample size. No comparison to a neurologic control group.
2004	Baldo, Delis, Wilkins & Shimamura Archives of Clinical Neuropsychology Both the frontal lesion and control groups were recruited from the same community	**D-KEFS TQT**	To compare the performance of patients with prefrontal cortex lesions on the D-KEFS TQT to healthy controls	Patients with frontal lesions performed worse on the D-KEFS TQT, asking significantly more questions than controls to identify target items across all 4 trials. Qualitatively, patients with frontal lesions tended to rely on ineffective and less highly concrete strategies, focusing more on concrete attributes (e.g. "Does it use gasoline?") and specific items (e.g. "Is it the banana?") that restricted them to narrow down the search. Poor performance on the D-KEFS TQT was strongly correlated with a fewer number of correct sorts on the D-KEFS ST (r = −0.75, p < 0.01).	Relatively small sample size. No comparison to a neurologic control group.
2001	Baldo, Shimamura, Delis, Kramer & Kaplan	**D-KEFS:** 1. VFT (Letter Fluency, Category Fluency, Category Switching)	To compare the performance of patients with focal frontal lesions on D-KEFS fluency	Compared with controls, patients with frontal lesions produced less designs on the DFT and fewer correct responses on the VFT. On design fluency, both the frontal and control groups showed similar pattern of	Small sample size.

| | Journal of the International Neuropsychological Society

No information about the sources from which the frontal lesion and control groups were recruited | 2. DFT (Filled Dots, Empty Dots, Dot Switching) | tests, VFT and DFT to healthy controls | performance across the 3 conditions, with participants from both groups performing worst in the switching condition. Across the 3 conditions on verbal fluency, patients with frontal lesions produced significantly fewer words in Letter Fluency when compared with controls. However, the frontal group did not appear to show a larger cost in the Category Switching. The frontal lesion group was not disproportionately impacted by the switching conditions of the DFT and VFT.

With regard to the lateralization effects, there was no significant difference between the left- and right-frontal patients in their performances on the DFT. However, patients with left frontal lesions performed worse than patients with right frontal lesions on the VFT, generating fewer items in the task. | No comparison to a neurologic control group. |

Neurodegenerative Disorders

Year	Authors, publication & source of participants recruitment	Measure(s) used to assess executive functions	Study aims	Summary of main findings	Limitations
2017	Gansler, Huey, Pan, Wasserman & Grafman

Cognitive Neurology

Participants with the required clinical diagnoses were recruited nationally via advertisement

No healthy control group | **DKEFS:**
1. VFT (Letter Fluency)
2. TT
3. ST

FrSBe (caregiver-rated) | To compare the performances on tests of executive function and dysexecutive behaviour among groups of patients with behavioural variant frontotemporal dementia (bvFTD), primary progressive aphasia (PPA) and corticobasal syndrome (CBS), | Among the D-KEFS measures, both the TT and ST scores were lower in the bvFTD group than in the PPA and CBS groups, but there was no significant difference between the 3 groups on the VFT score. In the FrSBe, patients with bvFTD performed significantly worse than the PPA and CBS groups.

The co-joint use of DKEFS and FrSBe measures correctly distinguished 89% of bvFTD from CBS patients whereas 93% of bvFTD from PPA patients respectively. | The selection of using the performance-based D-KEFS and the caregiver report-based FrSBe reduced the likelihood of good fit to establish a relationship. |

			To examine the incremental utility in the co-joint use of executive function and dysexecutive behaviour measures in differentiating the bvFTD group from the PPA and from the CBS group	Confirmatory factor analysis showed that executive function and dysexecutive behaviour were distinct but moderately correlated constructs (r= −0.48, p<0.01). Brain imaging analysis revealed that the caudal left dorsolateral prefrontal and lateral temporo-parietal cortices were distinctively associated with executive function measured by the D-KEFS, whereas the bilateral cingulate gyrus, right subcallosal and right anterior frontal cortex were distinctively associated with measured by the FrSBe. The neural correlates of executive function and dysexecutive behaviour overlapped substantially in the rostral areas of the lateral and dorsomedial prefrontal cortex bilaterally.	
2013	Kaiser et al Neuropsychologia Clinical groups were recruited from neurological clinics of the university No healthy control group	**DKEFS:** 1. Proverb Test 2. DFT (Filled Dots, Empty Dots, Dot Switching) 3. TT **WCST** **TMT** **Phonemic 'FAS' Fluency Stroop**	To compare patients with behavioural variant frontotemporal dementia (bvFTD) and those with Alzheimer's disease (AD) on the performance of D-KEFS Proverb Test To examine brain regions which are associated with proverb interpretation	Patients with bvFTD, in comparison with those with AD, performed significantly worse on the Proverb Test in terms of accuracy and interpretation. Specifically, bvFTD patients tended to demonstrate more concrete interpretation for the common proverbs whereas AD patients were more likely to respond with abstract interpretations. However, there was no significant group difference on the uncommon proverbs. Poor performance in proverb interpretation was significantly correlated with anterior temporal lobe region (r= 0.64, p= 0.01) when dementia diagnosis was controlled for.	Small sample size. Free inquiry condition was not included.
2013	Meoded et al Dementia and Geriatric Cognitive Disorders No information about the source from which the clinical groups and	**D-KEFS:** 1. TMT (Number-letter Switching) 2. VFT (Letter Fluency, Category Fluency, Category Switching) 3. ST (Free Sorting, Sort Recognition)	To compare the cognitive performance between patients with PLS and those suffering ALS To examine the relationship between deficits in performance	The PLS and ALS groups did not significantly differ in performance on any of the D-KEFS measures. The ALS group had higher proportion of patients meeting the criteria for cognitive impairment than the PLS group. ALS patients with cognitive impairment performed worse on multiple measures of the D-KEFS than those without cognitive impairment.	Small sample size. Sample was biased with high functioning patients.

	healthy controls were recruited	4. TQT 5. WCT 6. TT 7. PT **FrSBe** (caregiver-rated)	on the cognitive tests and imaging metrics	Significant associations were found between the D-KEFS performance and the diffusion metrics of white matter tracts. Factor analysis of the nine D-KEFS tasks yielded three significant factors: Fluency, Sorting and Trails. The Fluency and Sorting factors were associated with axial diffusivity of several white matter tracts, suggesting that the performance of these D-KEFS tasks required the integrity of tracts connecting the frontal lobes with different regions of the brain.	
2012	Clark, Schiehser, Weissberger, Salmon, Delis & Bondi Journal of the International Neuropsychological Society Participants were recruited from community via newspaper advertisements and flyers placed in senior centres and from a research centre for Alzheimer's disease No comparison group in this study	**D-KEFS:** 1. VFT (Category Switching) 2. DFT (Dots Switching) 3. TT 4. CWIT (Inhibition/ Switching) 5. TMT (Number-letter Switching)	To determine any specific changes in executive functions prior to the cognitive decline in older adults using the switching paradigms of the D-KEFS, and what particular aspects of executive functioning can best predict the decline	Among the non-demented older adults participated in the study, 15 of them exhibited a cognitive decline (decline group) in a 1-year period whereas the remaining 56 remained stable in cognitive functioning (no-decline group) as measured by DRS. In the year prior to the decline, the decline group performed significantly worse than the no-decline group on the CWIT and VFT switching conditions. Conversely, decliners and non-decliners performed comparably on the DF switching, TT spatial planning and TMT switching. Regression analysis indicated that the CWIT Inhibition/ Switching condition was the strongest predictor with the largest effect size, reliably distinguishing decline from no-decline outcome over 1-year period. Overall, the model correctly classified 73% of the sample with sensitivity and specificity rates of 80% and 71%.	Small sample size. No longitudinal followed up was pursued, so it was uncertain how many participants would develop further cognitive deterioration. Different procedure to classify groups was used in this study. Changes in executive functioning were assessed by only one to two measures of each task.

2009	Fine et al Journal of Clinical and Experimental Neuropsychology No information about the sources from which the clinical and control groups were recruited	**D-KEFS ST** (Free Sorting)	To examine relationships between lobar volumes of different brain regions and the performance on the D-KEFS Sorting Test in a group of mixed neurologically impaired patients	Among all six (bilateral frontal/ temporal/ parietal) lobar volumes, only left frontal lobar volume significantly predicted free sorting description score when the effects of potential moderators (e.g. demographic variables, MMSE score) and head size were controlled for. However, no significant relationships could be found between any of the lobar volumes and the total number of correct sorts. Further analysis indicated that the healthy controls scored significantly better on the free sorting description than the patient groups but there were no significant differences for any comparisons between patient groups. Similar pattern of group differences was also observed for the total number of correct sorts on the ST.	No comparison to a focal non-frontal control group. Groups were not matched with overall level of cognitive functioning dud to small sample size.
2009	Huey et al Neurology The clinical group was either self-referred or referred by outside neurologists The control group was recruited locally via advertisement and participants were paid for taking part in the study	**D-KEFS:** 1. TMT (all 5 conditions) 2. VFT (Letter Fluency, Category Fluency, Category Switching) 3. ST (Free Sorting) 4. TQT 5. TT	To compare D-KEFS performances between patients with bvFTD and patients with CBS To determine the associations between executive dysfunctions and brain regions involved in these pathologies	Discriminant analysis indicated that patients with bvFTD performed significantly worse than patients with CBS on the majority of the ST, TQT, TT and VFT. Of these, measures on TQT and rule violations performed best to distinguish the clinical diagnoses of CBS and bvFTD, on which the bvFTD group scored lower than the CBS group at greatest discrepancy. Patients with bvFTD, however, did better than patients with CBS on TMT Visual Scanning, and on two timed measures of the TT. Significant associations between D-KEFS performances and frontal regions of the brain were found. For instance, verbal fluency was associated with areas of left frontal perisylvian cortex, sorting was related to dorsolateral prefrontal cortex, and performance on TQT was correlated with left anterior frontal cortex.	The effects of general cognitive dysfunction and disease severity had not been controlled for.
2008	Carey et al	**D-KEFS TT**	To evaluate the discriminant ability of	When compared to healthy controls, both the FTD and AD groups demonstrated significantly lower total	Gender and dementia severity

	Neuropsychologia Both clinical and control groups were recruited from a memory and aging centre at university		rule monitoring to differentiate individuals with Frontotemporal Dementia (FTD) from Alzheimer's Disease (AD) To examine the neuroanatomical correlates of rule monitoring to determine the specificity to frontal changes at neuroanatomical level	achievement scores on the D-KEFS TT but there was no significant difference between the FTD and AD groups. In terms of rule violation, the FTD group committed significantly more errors than both the AD and control groups. Regression analysis revealed that the rule violation was a significant predictor of group membership for FTD versus AD, yielding an overall classification accuracy of 65%. Committing five or more rule violations on the task was an optimal cut-off point that demonstrated good specificity (80%) but low sensitivity (50%). Additionally, individuals who committed five or more rule violations were more than four times (odds ratio = 4.4) more likely to have a FTD diagnosis. With regard to the correlations between D-KEFS performances and brain regions, the total achievement score was positively associated with left frontal ($r = 0.5$, $p < 0.001$), right frontal ($r = 0.41$, $p < 0.01$), left parietal ($r = 0.35$, $p < 0.05$) and right parietal ($r = 0.39$, $p < 0.01$) brain volumes. In contrast, the rule violation was negatively correlated with left ($r = -0.39$, $p < 0.01$) and right ($r = -0.31$, $p < 0.05$) frontal volumes only. Bilateral frontal volumes still significantly predicted rule violations even after the effects of total achievement score and other brain areas were controlled for.	were not matched for group comparisons. Underlying mechanisms for the failure on the task such as processing speed or memory impairment were not controlled for. Lack of more specific imaging techniques to investigate the extent of brain atrophy among groups.
2008	Fine et al The American Journal of Geriatric Psychiatry No information about the source from which	**D-KEFS CWIT** (Colour Naming, Word Reading, Inhibition, Inhibition/ Switching)	To investigate the utility of the D-KEFS CWIT in performance of predicting cognitive decline over a year of period in a sample of older adults with APOE-e4 and without APOE-e4	Among the non-demented older adults participated in the study, 8 of them exhibited a cognitive decline (decline group) in a 1-year period whereas the remaining 11 did not demonstrate a decrease in cognitive functioning (stable group) as measured by DRS. In the year before their cognitive changes, the decline group performed significantly worse than the stable	Relatively small sample size. No longitudinal follow-up analysis was pursued, so it was uncertain how many participants

Year	Study	Test	Aim	Results	Limitations
	the participants were recruited No comparison group in this study			group only on the Inhibition/ Switching condition in terms of the completion time. CWIT discrepancy score (the contrast performance between the switching and fundamental conditions) was calculated for each participant. The decline group demonstrated a significantly higher mean discrepancy score than the stable group the year before their decline. Regression analysis showed that the discrepancy score before decline significantly predicted cognitive decline over the subsequent year. The model correctly classified 75% of participants into groups. However, the APOE status alone did not significantly predict the decline.	in the decline group would go on to develop AD.
2008	Nutter-Upham et al Archives of Clinical Neuropsychology Participants were recruited from flyers, newspaper advertisement, public lectures, and referrals from medical centre clinics	**D-KEFS VFT** (Letter Fluency, Category Fluency, Category Switching)	To examine the verbal fluency performance in older adults with amnestic MCI, older adults with cognitive complaints and healthy matched controls	When compared with healthy controls, participants with amnestic MCI scored significantly poorer in performance of verbal fluency on most measures (Letter, Category, Switching fluency and number of correct shifts between categories). Factor analysis of fluency tasks yielded two significant factors: Switching and Production. Discriminant function analysis was conducted to determine which factor best discriminated MCIs from controls. Results showed that the Switching factor correctly classified 24 of 33 healthy controls and 25 of 37 MCIs (overall classification accuracy of 75%) whereas the Production factor correctly classified 20 of 33 healthy controls and 22 of 37 MCIs (overall classification accuracy of 60%).	Limited generalisability of the findings due the homogeneity of the sample. The predictive ability of the verbal fluency task could not be examined in this study due to the cross-sectional nature of the data.
2007	Kramer et al Journal of the International	**D-KEFS DFT** (Empty Dots, Dot Switching)	To investigate the relationships between brain lobar volumes and set shifting performance on the D-KEFS DFT in a	Only left (r = 0.32, p < 0.001) and right (r = 0.30, p < 0.01) frontal volumes were significantly associated with the ability of cognitive set shifting after controlling for the effects of global cognition and the fundamental performance (Empty Dots). Such correlation between	The differentiation between FTD and AD was difficult. It was therefore possible that some

				frontal volumes and set shifting still remained significant even when working memory was controlled for. Additionally, left frontal volume was slightly more associated with switching ability than right frontal volume. In contrast, the performance on the switching condition showed no correlation with both parietal and temporal lobar volumes.	participants might have initially been misdiagnosed. Analysis of lobar volume did not offer sufficient information about specific frontal regions.
Neuropsychological Society Participants were recruited from the memory and aging centre at university			sample of patients with dementia and normal controls		
2007	Parmenter et al Journal of Clinical and Experimental Neuropsychology The MS group was recruited through referrals and from those who paid for neuro-psychological evaluation at an urban academic medical centre No information about the source from which the control group was recruited	**D-KEFS ST** (Free Sorting) **WCST**	To examine the performances of a large sample of patients with MS on the D-KEFS ST and WCST, relative to matched controls	Patients with MS performed significantly worse than controls on both D-KEFS ST and WCST. However, only the D-KEFS "Correct sorts", "Sort description" and "Sorting repetition" could significantly discriminate between the MS patients and controls when the effect of depressive symptomatology was controlled for. Significant correlations were found between measures of D-KEFS and WCST, with coefficients ranging from $r = -0.27$ to 0.54. Brain imaging analysis also showed that cognitive performances on both tests were modestly or strongly correlated with MRI brain indices. For instance, the D-KEFS "Correct sorts" was negatively associated with lesion volume ($r = -0.4$, $p < 0.05$) and positively associated with the extent of brain atrophy ($r = 0.58$, $p < 0.01$); the "Sort description" was negatively associated with lesion volume ($r = -0.44$, $p < 0.05$) and positively associated with brain atrophy ($r = -0.58$, $p < 0.01$); the "Sorting repetition" was positively associated with lesion volume ($r = -0.46$, $p < 0.05$) and negatively associated with brain atrophy ($r = -0.36$, $p < 0.01$). Performance on both tests was also correlated with patients' vocational disability. Regression analysis	Longitudinal changes in executive functioning could not be examined in this study due to its cross-sectional nature of data. Longer and more comprehensive measures to control for depressive symptomatology might be used in future study.

				showed that the D-KEFS description score was the best predictor, discriminating between employed and disabled MS patients.	
2005	Houston el al Journal of the International Neuropsychological Society All participants were recruited from community and the Alzheimer's Disease Research Centre at university through newspaper advertisement and community lectures	**D-KEFS:** 1. VFT (Letter Fluency, Category Fluency, Category Switching) 2. DFT (Filled Dots, Empty Dots, Dot Switching)	To examine any cognitive asymmetries in performance on verbal and visuospatial abilities in older adults with APOE-e4 on the D-KEFS VFT and DFT, relative to those without APOE-e4	In terms of mean raw scores on any subtests of the VFT and DFT, no statistical differences could be found between the APOE -e4 and non-e4 groups. Instead of comparing test mean scores, a standardised discrepancy score for the 2 measures was calculated for each participant. Discrepancy analysis revealed that approximately 38% of the participants with APOE-e4 exhibited an asymmetric performance (discrepancy score ≥ 1.5 SD) on the verbal and spatial fluency switching conditions, whereas only 14% of non-e4 participants demonstrated such cognitive discrepancies. Additionally, among those who showed an asymmetric profile, there was a slight trend for the APOE -e4 group to be more likely to display a "high spatial" asymmetric pattern, but not "high verbal" asymmetry, than the non-e4 group when the frequency of asymmetric cognitive profile type relative to the entire group was compared between APOE genotype groups. However, there was no significant group difference in the frequency of an asymmetric profile for any of the fundamental fluency tasks.	Individuals who had the greatest genetic risk factor for AD were not included in the APOE -e4 group. No longitudinal follow-up analysis was pursued, so it was uncertain how many participants in the decline group would go on to develop AD.
2005	Wetter et al Journal of Clinical and Experimental Neuropsychology Participants were recruited from a	**D-KEFS CWIT** (Colour Naming, Word Reading, Inhibition, Inhibition/ Switching)	To examine any subtle deficits in executive functioning in older adults with APOE-e4 on the D-KEFS CWIT, relative to those without APOE-e4	There was no significant difference between the APOE -e4 and non-e4 groups in their performances of total completion time on any of the CWIT conditions. However, the APOE-e4 group significantly committed more errors on the Inhibition/ Switching condition when compared to the non-e4 group. Those who have unusually high error rates were all in the APOE-e4 group. A significant association was found between the	No longitudinal follow-up analysis was pursued, so it was uncertain whether the APOE –e4 group had more people developing

				error rates in the Inhibition/ Switching condition and the global cognitive functioning as measured by the Dementia Rating Scale, but this correlation only existed for the e4 group (r = −0.47, p < 0.05), and not for the non-e4 group.	preclinical AD than the non-e4 group.
	research Centre for Alzheimer's Disease				

Epilepsy

Year	Authors, publication & source of participants recruitment	Measure(s) used in the study to assess executive functions	Study aims	Summary of main findings	Limitations
2017	Reyes et al Epilepsy & Behaviour TLE participants were recruited through referral from Epilepsy centres No information about the source from which the control group was recruited	**D-KEFS:** 1. CWIT (Inhibition/ Switching) 2. VFT (Category Switching) **TMT B**	To investigate the association between the frontostriatal network integrity, as reflected by the neurite density, and the performances on the D-KEFS measures in patients with TLE	Patients with TLE significantly performed worse than healthy controls on the D-KEFS Category Switching and TMT-B, but the two groups did not differ in the CWIT switching condition. A significant correlation between reduced neurite density within inferior frontostriatal tract and poorer performances on the CWIT Inhibition/ Switching was found only in patients with left-sided TLE.	The causality of executive deficits could not be inferred based on the correlational data. Other measures of executive functioning could be used to assess different domains.
2010	Luton, Burns & DeFilippis Archives of Clinical Neuropsychology Children with FLE were recruited from an on-going study at a	**D-KEFS:** 1. VFT (Category Fluency, Category Switching) 2. TMT (Number Sequencing, Letter Sequencing, Number-letter Switching) **BRIEF** (parent-rated)	To examine the executive functioning of children with FLE, with specific aim of investigating the effects of age of seizure onset on the measures of the executive function	Compared to normal controls, children with FLE demonstrated significantly greater difficulty on all measures of the D-KEFS and the BRIEF. Notably, children with early seizure onset performed significantly worse than controls on Category Fluency, Category Switching and Number-letter Switching. Conversely, children with later seizure onset performed comparably to those of controls on these tasks.	Comparisons were evaluated using qualitative methods, thus lacking in power for the analysis.

| | | | | | Higher risk of committing Type II error due to small sample size.

Inclusion of children with attention-deficit/ hyperactivity disorder in the FLE group. |
|---|---|---|---|---|---|
| 2009 | Pulsipher et al

Epilepsia

The epilepsy groups were recruited from paediatric neurology clinics at two large outpatient medical centres

Research participants' first degree cousins were recruited as healthy controls | **D-KEFS:**
1. ST (Free Sorting)
2. VFT (Category Switching)
3. CWIT (Inhibition)

BRIEF (parent-rated) | To determine the relationship between thalamofrontal volumetric abnormalities and executive function in children with recent onset juvenile myoclonic epilepsy

To compare the executive functioning of children with recent onset epilepsy to healthy controls using BRIEF | The epilepsy group performed significantly worse than controls on D-KEFS Inhibition and the BRIEF, with moderate to large effect sizes.

Regression analysis indicated that frontal and thalamic volumes were the best significant predictors of performance on all D-KEFS measures. For instance, bilateral thalamic volumes predicted performance on number of correct sorts; frontal tissues predicted performance on Category Switching and Inhibition. | The groups differed significantly in mean age.

Small sample size.

The potential effects of medication contributing to brain volumes and cognitive functioning were not determined. |
| 2008 | McDonald et al

The Clinical Neuropsychologist | **D-KEFS Proverb Test**
(Free Inquiry condition, Multiple Choice condition) | To compare the performance on D-KEFS Proverb Test between groups of patients with | Patients with FLE performed significantly impaired than controls on the Proverb Test in both response accuracy and abstractness but those with TLE did not differ from controls on the task. | Small sample size.

Other measures of verbal |

			FLE and TLE, relative to matched healthy controls To investigate seizure-related variables which are related poor performance	In patients with FLE, impaired performance was only associated with left-sided seizure focus whereas in patients with TLE, poor performance was related to early age of seizure onset as well as left-sided seizure focus. Using a performance 1.5 SD below the mean of the control group as the impairment cut-off, the free inquiry condition correctly classified 69% patients with FLE and TLE which demonstrated a high specificity (80%) and a modest sensitivity (59%), whereas the multiple choice condition demonstrated the diagnostic accuracy of 72% with sensitivity and specificity were of 55% and 75% respectively. However the diagnostic accuracy in the free inquiry condition significantly increased to 85% (sensitivity= 80%; specificity= 80%) when only left-sided FLE patients were included.	abstraction might be included in future study.
The two clinical groups were recruited from an epilepsy centre at university A matched group of healthy controls was obtained from the D-KEFS normative database					
2007	Parrish et al Developmental Medicine and Child Neurology The epilepsy group was recruited from paediatric neurology clinics of two large outpatient medical centres Research participants' first degree cousins were recruited as healthy controls	**D-KEFS:** 1. ST (Free Sorting) 2. VFT (Category Switching) 3. CWIT (Inhibition) **BRIEF** (parent-rated)	To compare the executive functioning of children with recent onset epilepsy to healthy controls using BRIEF and subtests of the D-KEFS To examine any correlation between the BRIEF and performance on the D-KEFS, and determine whether the BRIEF is a good predictor of D-KEFS performance	Children with recent onset epilepsy performed significantly worse than normal controls on the BRIEF and two of the three D-KEFS measures (ST and CWIT). The BRIEF Metacognition Index scale was significantly correlated with all three D-KEFS measures. The epilepsy group was split into two subgroups based on the BRIEF scores, the 'at-risk' group (score > 60) and the 'low risk' group (score < 60). The 'at-risk' group performed significantly worse than the 'low risk' group on all three D-KEFS measures.	The potential effects of taking antiepileptic medication was not controlled for the analysis.

2005a	McDonald et al Epilepsy & Behaviour The two clinical groups were recruited from an epilepsy centre at university A matched group of healthy controls was obtained from the D-KEFS normative database	**D-KEFS CWIT** (Colour Naming, Word Reading, Inhibition, Inhibition/ Switching)	To compare the performance on D-KEFS CWIT between groups of patients with FLE and TLE, relative to matched healthy controls	Patients with FLE performed significantly slower than the matched controls across all four conditions, with the FLE group showing noticeably larger impairment in the Inhibition and Inhibition/ Switching conditions. Notably, the FLE group also committed more errors than the control group in the Colour Naming and Inhibition/ Switching conditions. Conversely, the FLE and TLE groups did not significantly differ in their performance on the tests. However, subgroup analysis revealed that patients with left-hemisphere FLE were the most disadvantaged on both the Inhibition and Inhibition/ Switching conditions. No significant correlations were found between the seizures-related variables and CWIT performance in patients with FLE.	Sub-group analysis within the FLE group was limited by small sample size. Lack of information such as the side of the seizure focus or lesion location and such factors might have contributions to cognitive performances.
2005b	McDonald, Delis, Norman, Tecoma & Iragui-Madoz Journal of the International Neuropsychological Society The two clinical groups were recruited from an epilepsy centre at university A matched group of healthy controls was obtained from the D-	**D-KEFS TMT** (Visual Scanning, Number Sequencing, Letter Sequencing, Number-letter Switching, Motor Speed)	To examine any differences in performance of D-KEFS TMT between groups of patients with FLE and TLE, relative to matched healthy controls	Patients with FLE were disproportionately impaired on the Number-letter Switching condition relative to patients with TLE and healthy controls, whereas the three groups were indistinguishable on all the baseline conditions. In terms of accuracy, the FLE group committed more set-loss errors than the TLE and control groups. Conversely, the TLE and control groups performed comparably across all five conditions, including the switching condition. No significant correlations were found between the seizures-related variables and set-shifting performance in patients with FLE.	No comparison to a non-frontal control group. Sub-group analysis within the FLE group was limited by small sample size.

2005c	KEFS normative database				
	McDonald, Delis, Norman, Tecoma & Iragui Neuropsychology The two clinical groups were recruited from an epilepsy centre at university A matched group of healthy controls was obtained from the D-KEFS normative database	**D-KEFS DFT** (Filled Dots, Empty Dots, Dot Switching)	To examine the performance of patients with FLE on D-KEFS DFT, relative to patients with TLE and matched healthy controls To evaluate the ability of the DFT switching condition in discriminating between the FLE and TLE groups	Patients with FLE generated fewer accurate designs in the switching condition than both the TLE patients and healthy controls, whereas all three groups performed comparably on the two basic (Filled Dots and Empty Dots) conditions. Notably, the TLE and control groups did not differ in their performance in the switching condition. Subgroup analysis indicated that patients with left-side FLE were the most disadvantaged in terms of the number of correct designs generated in the switching condition than those with right-sided FLE and controls. Furthermore, committing more set-loss errors was also demonstrated in those with left-lesional and nonlesional FLE on the switching task. When using a performance 1.5 SD below the mean of the control group as the impairment cut-off, the switching condition correctly classified 72% patients with FLE and TLE which demonstrated a high specificity (90%) and a modest sensitivity (57%).	Sub-group analysis within the FLE group was limited by small sample size. Lack of information such as the side of the seizure focus or lesion location and such factors might have contributions to cognitive performances.

Note: D-KEFS test = Delis-Kaplan Executive Function System test; VFT = Verbal Fluency Test; WCIT = Word-Colour Interference Test; TT = Tower Test; ST = Sorting Test; TMT = Trail Making Test; DFT = Design Fluency Test; TQT = Twenty Questions Test; PT = Proverb Test; WCT = Word Context Test; FrSBe = Frontal System Behaviour Scale; BADS = Behavioural Assessment of the Dysexecutive Syndrome; WCST = Wisconsin Card Sorting Test; TMT A & B = Trail Making Test A & B; BRIEF = The Behaviour Rating Inventory of Executive Function; MMSE = Mini-Mental State Examination; DRS = Dementia Rating Scale; CVLT-II = California Verbal Learning Test-Second Edition; BNT = Boston Naming Test; WAIS-III = Wechsler Adult Intelligence Scale-Third Edition; WMS-III = Wechsler Memory Scale-Third Edition; SD = Standard Deviation

<u>Screening Questionnaire</u>

The purpose of this questionnaire is to allow our research team obtaining some clinical information about you such as your background information, injury history or diagnosis which are relevant to our study.

You will be asked to provide the following information:

<u>Personal information</u>

Name:

Age:

Sex:

Ethnic/ Race background:

Primary language:

Education:

Hand used for writing: [   ] Right   [   ] Left

Hearing: [   ] Normal   [   ] Impaired

Vision: [   ] Normal   [   ] Impaired

<u>Medical history</u>

Diagnosis of neurological illnesses (e.g. Alzheimer's, Autism, Epilepsy, Parkinson's):

[   ] No   [   ] Yes, details _____

History of sensory/ motor/ cognitive/ language deficits

[   ] No   [   ] Yes, details _____

Psychiatric illness: [   ] No   [   ] Yes, details _____

Alcohol/ Substance misuse: [   ] No   [   ] Yes, details _____

Previous history of sustained brain injury: [   ] No   [   ] Yes, details (Experience of unconsciousness, hospitalisation)

_____

_____

The following part will be completed by the orthopaedic consultant with your permission.

I, the orthopaedic consultant of _____ hereby verify the above information is correct**.** Please tick as indicated.

[   ] Yes, the above information is correct

[   ] Additional information may be needed

_____

_____

Orthopaedic diagnosis: _____

Current Medication: _____

_____

_____

Volume I: Appendix IV – Diagnosis and classification of TBI

Mild uncomplicated TBI was defined in this study as any TBI where PTA was < 24hrs, GCS 13-15 and LoC duration < 30 minutes and where there were no brain imaging abnormalities.

Mild complicated TBI was defined in this study as any TBI where PTA was < 24hrs, GCS 13-15 and LoC duration < 30 minutes but where there was the presence of imaging abnormalities such as depressed skull fracture, evidence of bleeding or contusions.

Moderate injury severity included persons with a GCS of between 9 and 12, a PTA duration of greater than 24 hours and less than 7 days and a LOC of less than 6 hours.

Severe injury severity included persons with a GCS of 3 to 8, PTA duration of greater than 7 days and a LOC of greater than 6 hours. The PTA classifications were used as per American Congress Rehabilitation Medicine definitions (i.e. mild <24hrs, moderate < 1 week, severe > 1 week).

University Hospitals **NHS**
Birmingham
NHS Foundation Trust

PARTICIPANT INFORMATION SHEET

Title of the research study: Investigating the validity of the D-KEFS test as a neuropsychological assessment tool for executive functions in Traumatic Brain Injury (TBI) in the UK

We would like to invite you to take part in a research study. Before you decide whether or not you wish to take part, it is important for you to understand why the research is being carried out and discuss any questions you may have with the researcher. Please take time to read the following information carefully and discuss any questions you may have with the researcher. Please ask if there is anything that is not clear to you or if you would like any further information.

What is the purpose of the study?

The following research study will be completed by trainee from the DClinPsy Clinical Psychology course at the University of Birmingham. The primary aim of our research is to investigate the cognitive effects (e.g. planning, attention, flexibility of thinking) of traumatic brain injury (head injury). In order to address this we require a comparison group of participants who have not sustained a traumatic brain injury but have had a non-head related trauma. This allows us to account for both lifestyle and background factors that might make people more likely to experience trauma as well as the more general effects of having been through a traumatic event including stress, pain and medication. We have excluded people with spinal injuries, neck injuries and head/facial trauma to ensure that people with head injuries are not accidentally included in the comparison group.

The Delis-Kaplan Executive Function System (D-KEFS) provides a comprehensive assessment of higher-level cognitive functions following Traumatic Brain Injury (TBI). Ultimately, we would like to ensure that this tool is sensitive to evaluate "executive functions" such as attention, planning or self-organising ability following TBI. This is useful to allow for planning of treatment needs and rehabilitation programmes of people with head injury. If you take part in this study, you will be asked to complete some assessment tests in relation to executive functions as well as additional measures of reading, mental speed and concentration which allow us to look at the specific causes of any poor scores on the tests in the head injury group.

Why have I been chosen?
You are being invited to participate because you have experienced a traumatic injury that did not include significant injury to your head and are aged between18 and 89.

Do I have to take part?
No. Your involvement in this study is voluntary. If you decide to take part, you are still free to withdraw at any time without needing to give reason for this. Any decision you make to withdraw, or a decision not to take part at all, will have no effect on the standard of healthcare you receive now or in the future.

What will happen to me if I take part?
The researcher will meet with you at the NHS site where you receive your current medical treatment to carry out some assessments with you. These will consist of around 10 neuropsychological tests that are routinely used to assess cognitive functioning. All of the tests will be carried out either using a pen and paper or a laptop computer. There will be no physical examinations or medical procedures.

The tests will take around 60 minutes to complete in total and you will be provided with breaks as required. We also encourage you to ask the researcher for breaks at your convenience. Alternatively, it is possible for testing to take place over 2 sessions or on another occasion if you are too fatigued on the day; again this is at your convenience.

We will not access your medical records directly. However, we will ask you to fill in a brief questionnaire to allow us to have some additional background information. We will also, with your permission, ask your orthopaedic consultant to fill in the section detailing the nature of your orthopaedic injuries, your current medication and verifying your medical history.

Your scores from the tests will be entered into a spread sheet for analysis where you will only be identified by a number. This means that your data will always remain anonymous. The informed consent form that you will sign will be stored in a locked cupboard at the University Hospitals Birmingham along with all other participants' consent forms. You will also receive a copy of the information sheet and signed consent form, with another copy of these documents being kept in a locked cabinet at the University Hospitals Birmingham.

Are there any risks of taking part?
We do not expect that any part of this study will cause harm to anyone taking part. Some people may become tired during testing and in this case we encourage you to inform the researcher so that you are able to take a break. Other people may become frustrated whilst taking part in some of the assessments. We would like to emphasise that all of the tests you are going to complete have been designed to assess a wide range of functioning and there will be some aspects of these tests that are purposefully very difficult to complete. Consequently many people in the normal population would be expected to fail some items. If we consider there are any concerns regarding your neurological condition, your scores from this study will be fed back to your current clinical team with your permission. You will be reviewed by your responsible medical officer and may signpost to the appropriate care pathway but would, by definition, be excluded from the current study.

If at any point during or after testing you experience distress and would like to discuss it with someone, the contact details of the researcher and the collaborator at your NHS site are provided below. Equally, your clinical team will be more than willing to discuss any issues that you may have.

Are there any benefits of taking part?
On a personal level, there will be no benefit to you as a result of participating in the study. However, your participation in this research will provide data that will be extremely beneficial in assisting to further research into the validity of the D-KEFS in patients with TBI in the UK, which is an important area of research. Such investigation will be beneficial in the development of a more sensitive neuropsychological assessment tools to assist the diagnosis of people with traumatic brain injury and their rehabilitation needs.

What will happen when the research study stops?
The data will be entered into a database and analysed together with data from other participants who took part in the study. The results will be published in journal articles, however, your identity or involvement in the study will never be revealed. It will be possible for you to see the results of the study when it is finished.

If you would like an individual feedback, a summary of the whole findings at the end of the study will be given as per your request only. This can be provided by email or through some other convenient means of communication. However, it is anticipated that the findings will be presented publically when the study has been published in academic journals.

Will I remain anonymous?
All contributions you make towards this research study will be anonymous. Your data will be stored securely in a spread sheet that is only accessible by members of the research team. Your scores will be identified by an individual code and your name will not be used at any time other than on the written consent form, which will be stored in a locked cupboard at the University Hospitals Birmingham. Should a participant withdraw from the study at any time, data will also be retained safely and securely in an anonymised form.

Who has reviewed the study?
All research in the NHS is examined by an independent group of people called a Research Ethics Committee. Their role is to protect your safety, rights, wellbeing and dignity.

Who do I contact for further information?
If you require any further information then please contact Chan Yin Ming, Frances (Chief Investigator of the study) at ymc404@student.bham.ac.uk, or Dr David Hacker (Consultant Neuropsychologist) at Queen Elizabeth Hospital Birmingham.

Queen Elizabeth Hospital Birmingham:
Dr David Hacker- _____ , 01213712000 ext 16870

Alternatively, if you would prefer to seek advice from an individual who is independent of the research study please use the following contact information:

Patient Advice and Liaison Service (PALS)- pals@uhb.nhs.uk; 01213713280 (Queen Elizabeth Hospital Birmingham; B15 2WB)

If you are unhappy at any point during your involvement in this research study then feel free to contact the Principal Investigator with any concerns:

Dr David Hacker at ⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛⬛, 01213712000 ext 16870

If you would like to express your interest in participating in this study then a member of your clinical team will inform the research team and we will contact you from there.

Thank you for reading this.