Effort Test Results: The Effect of Informed Consent in a Clinical Sample

Thesis submitted to the University of Leicester,

Faculty of Medicine and Biological Sciences,

School of Psychology, for the degree of

Doctorate in Clinical Psychology

by

Alice Nicholls BSc (Hons)

Department of Psychology - Clinical Section

University of Leicester

2012

Declaration

I confirm that the literature review, research report and critical appraisal contained within this thesis are my own work and have not been submitted for any other degree or to any other institution.

Effort Test Results: The Effect of Informed Consent in a Clinical Sample

Alice Nicholls

Thesis Abstract

Effort or Symptom Validity Tests (SVTs) are used during neuropsychological assessment to assess for negative response bias. SVT failure can be used as evidence that other test results are invalid and to support a diagnosis of malingering. The positive predictive accuracy of a SVT is dependent on its sensitivity, specificity and the base rate of malingering within the population sampled. The British Psychological Society (BPS, 2009) advises that all clinical patients should be assessed for effort using a SVT. However, there is no available data on the likely base rates of malingering within a UK clinical sample. Furthermore, despite test manual instructions, the BPS also advises that examinees should be informed they will be assessed for effort, potentially invalidating test results.

A systematic literature review was conducted to ascertain what is currently known about the base rates of malingering. Studies were only included if they enabled the application of the Slick, Sherman and Iverson (1999) criteria for definite or probable malingering to their sample. Four North American Studies yielded 503 litigating, traumatic brain injured participants of which 24.55% were identified as either probably or definitely malingering. This figure was significantly lower than previous estimates, which have suggested the base rate of malingering may be as high as 40% (Larabee, 2003).

In order to investigate whether informing people presenting for a neuropsychological assessment that they would be tested for effort affects their SVT results a multi-site experimental design was employed. Participants were randomly assigned to either informed or uninformed conditions and administered a battery of neuropsychological tests including the Test of Memory Malingering (Tombaugh, 1996). Practical difficulties resulted in small sample size and insufficient statistical power to either accept or reject the null hypothesis. Further data collection, research opportunities and clinical implications are discussed.

Acknowledgements

I would like to extend my thanks to my academic supervisor Professor Michael Wang who has been inspirational, supportive and, somehow, remained calm though out the research process.

This research was dependant on the support of collaborating clinicians, thanks go to everyone who contributed to data collection, including, but not limited to: Professor Michael Wang, Dr Charlotte Fuller, Ms Jeanette Forster, Dr Louise Birkett-Swan and Dr Declan McNicholl. The research would have been nothing without its participants and I would like to thank everyone who agreed to participate in the study.

I am extremely grateful to Penny, Pamela and Carl at 104 Regent Road for helping me to locate resources, forwarding important documents and locating course staff during my various personal and research crises. I am also incredibly lucky to have had three very understanding clinical placement supervisors. My thanks go to Dr Sharon Lord, Dr Caroline Knight and Dr Inga Stewart for being nothing other than supportive throughout the research process.

Finally, this thesis could have never been written without the support of my friends and family. Special thanks go to Annie, Odin and Bronwyn for their refusal to let me forget the important things in life. And to Dan; thank you for following me across the country, enduring the last three years, and sharing my world with me. This would not have been possible without you. We can go and 'get on with living' now.

Word Count

	Main Text Only	With Tables and References
Literature Review	6803	8521
Research Paper	10419	12105
Critical Appraisal	2882	3028
Non- Mandatory Appendices		2701
Total	**20104**	**26355**

Contents

The Base Rates of Malingering in Litigating TBI Samples: A Critical Review

Critical Appraisal

List of Tables

List of Figures

Appendices

The Base Rates of Malingering in Litigating TBI Samples: A Critical Review

The Base Rates of Malingering in Litigating TBI Samples: A Critical Review

Alice Nicholls

Abstract

Introduction: Post Traumatic Brain Injury (TBI) neuropsychological assessment is only accurate if examinees deploy their best effort and respond honestly. Malingering is defined as when examinees have an external incentive to mislead the examiner and make a conscious effort to perform poorly. Symptom validity tests are used to identify either a lack of effort or effort to perform poorly. The positive predictive accuracy of symptom validity tests vary as a function of malingering base rates within the sampled population. Estimates within the literature suggest that base rates of malingering could be as high as 40% (Larabee, 2003). However, no previous systematic reviews of the base rate data for malingering within TBI samples were found.

Method: A systematic search of the major databases for studies where TBI samples were screened for malingering as defined by Slick, Sherman and Iverson (1999) was conducted.

Results: Four North American Studies yielded 503 participants of which 24.55% were identified as either probably or definitely malingering. Methodological concerns suggest that this figure may be an overestimation and should not be applied outside of North America.

Discussion: The findings suggest that the base rate of malingering in TBI samples is significantly lower than previously estimated (Larabee, 2003) and is probably lower still outside of North America. Therefore, it is likely that the positive predictive accuracy of symptom validity tests is compromised, potentially resulting in a high rate of false positive test results. Clinical and research implications are discussed.

1. Introduction

1.1. Effort Issues in Neuropsychological Assessment

The term Traumatic Brain Injury (TBI) refers to any damage to the brain caused as a result of an external force. Such injuries include, but are not limited to; blows to the head caused by slips, trips and falls; road traffic accidents and gunshot wounds (Lezak, 2004).

Post TBI neuropsychological testing may be requested in order to gauge the extent of functional loss, as a diagnostic tool and to aid in rehabilitation planning. In some cases it is used to support claims for compensation, fitness to work or state benefits. Neuropsychological tests of cognitive functioning are only able to provide accurate assessment of current functioning if the patient deploys their full effort (Lezak, 2004). For this reason many test administration manuals instruct the assessor to remind the patient to "try their best" (e.g. Wechsler Adult Intelligence Scale; Wechsler, 1997, Behavioural Assessment of the Dysexecutive Syndrome; Wilson, Alderman & Burgess, 1996).

Despite encouragement from assessors some patients fail to deploy their best effort during cognitive assessment. If a patient intentionally misleads the assessor for reasons such as the psychological need to assume a sick or disabled role and to receive attention or social reinforcement it is referred to as a factitious disorder (APA, 2000). However, when a patient intentionally misleads the assessor as to their level of disability or disorder in order to gain financial compensation or medication, avoid work or criminal prosecution, it is referred to as malingering (APA, 2000).

1.2. Malingering

The DSM-IV (APA, 2000) refers to 'malingering' not as a mental disorder but as an additional condition that may be the focus of clinical attention; it represents a conscious decision and is not necessarily stable across time, person or situation (Drob, Meehan & Waxman, 2009). Slick, Sherman and Iverson (1999) proposed diagnostic criteria for malingered neurocognitive dysfunction whereby 'definite' malingering is defined by the presence of a substantial external incentive and definite negative response bias not accounted for by psychiatric, neurological or developmental conditions. It should be noted that a diagnosis of malingering can be made in the presence of genuine disability: Lipman (1962) made a distinction between four types of malingering; 1) Invention, where no genuine symptoms are present; 2) Perseveration, where the patient pretends that genuine symptoms have persisted after their resolution; 3) Exaggeration, where the patient has genuine symptoms but exaggerates their severity; 4) Transference, where the patient has genuine symptoms but fraudulently attributes them to a specific cause.

1.2.1. External Incentives

Both the APA (2000) and Slick, Sherman and Iverson's (1999) definitions of malingering include the presence of a 'substantial external incentive' as a necessary condition for diagnosis. Such incentives include money; evading prosecution or imprisonment; gaining drugs and avoiding work or military service. While there is a growing body of research investigating malingering in criminal defendants undergoing competency to stand trial assessments (e.g. Ardolf, 2007) most published research has investigated malingering in the context of substantial monetary incentives such as personal injury compensation or disability benefits.

1.2.2. Detection of Malingering

Historically there has been widespread confidence among clinicians that their specialised training is, in and of itself, sufficient for judging whether or not a client is malingering (Dawes, Faust & Meehl, 1989). However a recent meta analysis of 193 studies showed that psychologists were only slightly more accurate in their detection of deception than a sample of student researchers (62% versus 54% respectively; Aamodt & Custer, 2006). However, 29% of surveyed UK neuropsychologists commented that they believed Symptom Validity Test (SVT) use was unnecessary as malingering would be obvious from client presentation and or test results (McCarter, Walton, Brooks & Powell, 2009).

It is common for judgements regarding symptom validity to be made on the basis of unusual or implausible results on a battery of tests designed to assess cognitive functioning (Lezak, 2004). Many tests of cognitive functioning have developed scales that may indicate exaggeration or response invalidity. For example the Rarely Missed Index (RMI; Ord, Greve & Bianchini, 2008) of the Logical Memory subtest of the Wechsler Memory Scale (LM; WMS-III; Wechsler, 1997) identifies responses that are unlikely even in patients with extreme levels of impairment. Despite continuing development, the sensitivity and specificity of response validity measures within neuropsychological tests are generally inferior to those of tests specifically designed for this purpose (Van Gorp et al., 1999). As such the British Psychological Society (BPS; 2009) only endorses their use alongside tests specifically designed to test symptom validity.

1.3. Symptom Validity Tests

Most symptom validity tests for neurocognitive impairment employ a forced choice paradigm; examinees are exposed to pictures, numbers or words in a learning trial and then asked to choose between one target item and one foil in a recognition trial (Babikian & Boone, 2007). As the examinee has a 50% chance of selecting the correct item, significantly below chance performance suggests an effort to mislead the examiner. In the presence of an external incentive and the absence of psychiatric or neurological reasons for such poor performance, an examinee scoring significantly below chance (Binomial test, $p<0.05$) would meet the Slick, Sherman and Iverson (1999) criteria for 'definite' malingering. Forced choice SVTs are typically very easy to do well on and respondents with severe cognitive impairments tend to score well above chance (e.g. Tombaugh, 1996). Therefore, the examinee's effort is called into question when their score falls below the 'floor' performance of a normative severe head injury sample. Such a floor can be as high as 90% correct (e.g. Tombaugh, 1995) and failure at below-average rather than below-chance performance should be interpreted cautiously (Slick, Sherman and Iverson, 1999) as only partial fulfilment of the criteria for 'probable' rather than 'definite' malingering.

1.4. The Importance of Base Rates

The diagnostic validity of a Symptom Validity Test (SVT) depends on its ability to accurately detect malingering (sensitivity) and correctly classify people who are not malingering (specificity). The Positive Predictive Accuracy (PPA) of the test is the proportion of individuals identified as malingerers by the test who are actually malingering. The Negative Predictive Accuracy (NPA) is the proportion of individuals identified as honest by the test who are actually honest. In SVTs, PPA is arguably the

most important validity indicator (e.g. Rogers, 1997; Rosenfeld, Sands & Van Gorp, 2000) due to the potential damage a false diagnosis of 'malingering' could cause (e.g. exclusion from services, criminal prosecution and loss of financial benefits).

The base rate of any condition affects the PPA of its respective diagnostic test. For example, if the base rate of malingering was 50% and a test has a sensitivity of .90 and specificity of .95 in a sample of 200 the test would accurately classify 90 and wrongly classify 5 as malingerers (PPA= .95). Therefore, the lower the base rate of malingering is, the less accurate SVTs become due to a drop in PPA (BPS, 2009).

In a recent meta-analysis (Sollman & Berry, 2011) of the sensitivity and specificity of five commonly used memory Symptom Validity Tests (Test of Memory Malingering, Tombaugh, 1997; Letter Memory Test, Inman et al., 1998; Medical Symptom Validity Test, Green, 2004; Word Memory Test, Green 2003; Victoria Symptom Validity Test, Slick, Hopp, Strauss & Spellacy, 1996) the mean sensitivity and specificity across these tests were found to be 0.69 and 0.90 respectively. In a hypothetical sample of 200 with a malingering base rate of 50% the proportion of those correctly identified as malingering (PPA) would be 87%. However, if the base rate of malingering was 10% the PPA would drop to 43% meaning that a positive result on the SVT would be more likely a false positive than a true positive.

1.5. Base Rate Estimates

Malingering base rates are difficult to estimate due to the fact that malingerers are unlikely to admit to such behaviour (BPS, 2009). Base rates are likely to differ in relation to external incentives, the nature of the injury, socioeconomic status and the culture surrounding litigation in any given population (BPS, 2009). Malingering cannot be ruled in or out on the basis of neuroimaging findings (Lezak, 2004). Therefore,

previous base rate research has employed various criteria for identifying the likely base rates within different populations. Mittenberg, Patton, Canyock and Condit (2002) surveyed members of the American Board of Clinical Neuropsychologists. The returned surveys suggested that the Neuropsychologists had (at least) suspected malingering in 39% of the Mild Head Injury cases assessed. However, while the neuropsychologists had specified reasons for suspecting malingering (such as improbable test results) there was no evidence of strict adherence to diagnostic criteria for malingering (e.g. Slick, Sherman & Iverson, 1999). While Mittenberg et al.'s (2002) survey provides useful information regarding the proportions of patients that neuropsychologists suspect are malingering; it does not provide reliable information regarding the base rates of malingering in clinical populations.

1.6. Previous Reviews

Larabee (2003) is widely cited as having performed a literature review on the base rates of malingering in neuropsychological settings (e.g. BPS, 2009; Sollman & Berry, 2011). Within the introduction to an empirical research paper Larabee (2003) identified 11 studies providing malingering base rate information in neuropsychological settings. Without reference to a search strategy or inclusion criteria Larabee (2003) summed the participants from the 11 studies, yielding 1363 participants, of whom 548 were identified as malingering. Larabee (2003) concluded that the base rate of malingering was approximately 40% (range 15-64%) and notes that this figure is congruent with Mittenberg et al.'s (2002) figures.

On examination of Larabee's selected studies (Binder & Kelly, 1996; Frederick, Safaty, Johnston & Powell, 1994; Greiffenstein, Baker & Gola, 1994; Grote et al., 2000; Heaton, Smith, Lehman & Vogt, 1978; Meyers & Volbrecht, 1998; Millis, 1992; Millis,

Putnam, Adams & Ricker, 1995; Rohling, Green, Allen & Iverson, 2002; Trueblood & Schmidt, 1993; Youngjohn, Burrows & Erdal, 1995) methodological issues preventing a reliable estimate of base rates were detected. In two studies, the participants had no external incentive to malinger or incentives to malinger were not reported (Binder & Kelly, 1996; Greiffenstein, Baker & Gola, 1994 respectively). Several studies (Frederick, Safarty, Johnston & Powel, 1994; Grote et al., 2000; Heaton, Smith, Lehman & Vogt, 1978; Meyers & Volbrecht, 1998; Millis, Putnam, Adams & Ricker, 1995; Rohling, Green, Allen & Iverson, 2002) had reported biased responding rather than malingering (e.g. below average performance on one symptom validity test). These studies either did not employ sufficient measures or did not provide enough information to apply the Slick Sherman and Iverson (1999) criteria for probable or definite malingering to their sample. One study's sample (Millis, 1992) was preselected on the grounds of suspicious self-report of symptoms (i.e. not able to return to work following a mild head injury) probably inflating the base rate of malingering within this sample.

The majority of the studies reported by Larabee (2003) were not intended to provide base rate data for malingering and were addressing a different research question. For example, most of the studies set out to validate a measure of symptom validity (Binder & Kelly, 1996; Millis, 1992; Frederick, Safarty, Johnston & Powell, 1994; Greiffenstein, Baker & Gola; Meyer & Volbrect, 1998; Millis et al., 1995; Grote et al., 2000) or to exclude people with suspect effort for the purposes of examining a sample without compromised effort (Rohling, Green, Allen & Iverson, 2002). Indeed, Larabee (2003) did not purport to be providing a systematic review of the literature and was providing a brief estimate of base rates for the purposes of his own study. In the absence of any, more systematic review, many publications have relied on Larabee's review without questioning its reliability and, therefore, demonstrating the need for a

thorough, systematic review of the malingered neurocognitive impairment base rate data.

1.7. Aims of the Present Review

A diagnosis of malingered neurocognitive dysfunction can have severe ramifications for the individual, including denial of services, treatment and benefits. Such diagnoses should only be made where necessary and when there is no reasonable doubt that the individual is malingering. If accurate, Symptom Validity Tests (SVTs), when used in conjunction with the Slick, Sherman and Iverson (1999) criteria, may be a useful tool to aide diagnostic decision making. Without reliable estimates of the malingering base rates within samples reporting neurocognitive impairment the Positive Predictive Accuracy (PPA) of SVTs remains uncertain. Therefore, the present review aims to identify reliable information regarding the base rates of probable and definite malingering (Slick, Sherman & Iverson, 1999) within Traumatic Brain Injury samples.

2. Method

An initial search of Abstracts of Reviews of Effects (DARE) and The Cochrane Database of Systematic Reviews (CDSR) revealed a lack of systematic reviews of the malingering in brain injury base rate data.

A systematic review of the research relating to base rates of malingering in patients reporting acquired brain injuries was conducted. Databases were selected in order to identify articles from the fields of Psychology, Medicine, Science, Law and the Social Sciences (PsychInfo, MedLine, Science Direct, Web of Knowledge, Criminal Justice Abstracts and Applied Social Sciences Index and Abstracts). The following search string was entered into each data base:

> *((Malingering OR Maling* OR Faking OR Response Bias OR Susp* Effort OR Simulation OR Symptom Exaggeration OR Incomplete Effort OR Effort) AND (Base rate* OR Frequency) AND (Brain Injury OR Head Injury OR ABI OR TBI OR Post Concussion Syndrome OR PCS OR Cognitive Impairment))*

Search results were limited to books and articles published in the English language where the search terms appeared in the abstract or title of the article. Where possible search results were further refined by the exclusion of obviously irrelevant subjects areas (i.e. Gastroenterology and Artificial Intelligence). Where this was not possible results were sorted by relevance and the abstracts were reviewed. When results were sorted by relevance and the number of articles returned was over 50 the finding of 20 consecutive, irrelevant articles was deemed to end the search. When the relevance of the article was not clear from the abstract the whole paper was ordered and reviewed in its entirety. The search was deemed exhaustive when new database searches failed to detect any new articles.

Articles were electronically retrieved or requested from the library if their abstracts suggested they might have collected information on the base rates of malingering in an acquired brain injury population. The reference lists of such articles were searched to discover further potentially relevant articles. These two methods identified a total of 45 articles suitable for further investigation.

2.1. Inclusion Criteria

The 45 shortlisted articles were subjected to the following inclusion criteria:

1) The study investigates an Acquired Brain Injury (ABI) population
2) The study examines the base rate of malingering within an ABI sample
3) The study examines ABI patients with external incentives to malinger
4) The study identifies 'definite' and 'probable' malingering according to stringent application of the Slick, Sherman and Iverson (1999) criteria or provides enough data to allow application of the criteria (See table 1).

Table 1. The Slick Sherman and Iverson (1999) Criteria for Malingering

Criteria		Description
A	Presence of substantial external incentive	
B	Evidence from neuropsychological testing	
	1	Below chance performance ($p<0.05$) on one or more forced-choice test
	2	Failure on a well validated measure of symptom exaggeration or fabrication
	3	Neuropsychological test performance is inconsistent with known patterns of brain functioning
	4	Discrepancy between test data and observed behaviour

	5	Discrepancy between test data and reliable collateral reports
	6	Discrepancy between test data and documented background history
C	Evidence from self-report	
	1	Self-report history is discrepant with documented history
	2	Self-reported symptoms are discrepant with known patterns of brain functioning
	3	Self-reported symptoms are discrepant with behavioural observations
	4	Self-reported symptoms are discrepant with information obtained from collateral informants
	5	Evidence of exaggerated or fabricated psychological dysfunction (e.g. validity scales in self-report measures of psychological adjustment)
D	Behaviours meeting criteria from groups B or C are not fully accounted for by psychiatric, Neurological or Developmental factors.	

Diagnosis	Necessary criteria
Definite Malingering	A B1 D
Probable Malingering	A 2 (B2-B6) Or 1 (B2-B6) and 1(C1-C5) D

2.2. Exclusion Criteria

1) The study examines ABI patients with Co-morbid psychiatric diagnoses

2) The study's sample includes a heterogenous mix of diagnostic categories within ABI and does not provide separate malingering statistics for each diagnostic category
3) The study's sample has already been subjected to selection on the basis of suspicion of malingering
4) The study defines malingering as symptom validity test failure without reference to the other Slick, Sherman and Iverson (1999) criteria
5) The rate of malingering within the sample has been published elsewhere in a more relevant article

2.3. Data Extraction

Following application of the inclusion and exclusion criteria eight studies were deemed relevant to the current review. The following information was collated onto data extraction forms:

1) Authors
2) Date
3) Location
4) Type of ABI within Sample
5) Time (months) between injury and testing
6) Size of sample with both ABI and external incentive to perform poorly
7) Mean age within sample
8) Percentage of females within sample
9) Sampling method
10) Type of external incentive
11) Name of Forced-choice test

12) Percentage of sample meeting the Slick criteria for definite malingering

13) Percentage of sample meeting the Slick criteria for probable malingering

14) Method for identifying 'probable' malingering.

15) Purpose of identifying malingering within sample.

Where studies had included participants without an external incentive to malinger the studies were retained but participants without incentives were removed to reflect the number of participants with an external incentive to malinger. Several studies had included participants who scored significantly below chance on forced choice measures in 'probable malingerer' groups: these participants were removed from the 'probable' groups and where necessary placed in a separate 'definite malingerers' group. This ensured that participants in the 'definite' groups were independent of participants in the 'probable' groups. One study (Trueblood & Schmidt, 1993) had allocated eight participants to a 'questionable validity group' based on one highly improbable test result (e.g. zero grip strength): of these eight participants, six went on to fail two tests of symptom validity and were included in the 'probable group' for the purposes of the present review.

Three of the selected studies were published within a two-year period and shared authors (Greve & Bianchini, 2006; Greve, Bianchini, Love, Brennan, & Heinly, 2006; Curtis, Thompson, Greve & Bianchini, 2008). While these studies did not state they had used previously published sample data it seemed likely the samples might have shared participants. Communication with the named author (Dr Greve) confirmed that these samples were not independent, therefore the study with the largest number of participants was retained (Greve & Bianchini, 2006) and the remaining articles were excluded. Five articles remained relevant for inclusion in the analysis.

3. Results

Application of the inclusion criteria and removal of duplicated data sets yielded five studies suitable for further review. One study (Greve & Bianchini, 2006) reported results from two data sets therefore six separate data sets were available for review.

3.1. Participants

Across the six data sets 602 participants met the inclusion criteria of the present review. The majority of participants were male (64%) with an average age of 37, all spoke English as a first language and were resident in either North America or Australia.

3.1.1. Brain injury severity

Four of the data sets (Youngjohn, Burrows & Erdal, 1995; Greve & Bianchini, 2006 set 'a'; Langeluddeck & Lucas, 2004; Binder, 1993) examined the frequency of malingering in a Mild Traumatic Brain Injury (MTBI) sample. All studies defined MTBI according to the criteria set by the Mild Traumatic Brain Injury Committee of the Head Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation Medicine (1993):

1) Post Traumatic Amnesia (PTA) not greater than 24 hours
2) After 30 minutes, an initial Glasgow Coma Scale (GCS) of 13-15
3) Loss of Consciousness (if any) of less than 30 minutes

In addition, all MTBI participants had, or reported that they had, received a blow to the head and were excluded if they had positive neuro-radiological findings or focal neurological signs. In addition to the MTBI criteria, Youngjohn, Burrows and Erdal (1995) required participants to be reporting symptoms of persistent Post-Concussion

Syndrome (PCS) six months after their MTBI. They defined PCS in patients reporting any of the following symptoms: Memory impairment, headache, dizziness, concentration difficulties, blurred vision, photophobia, tinnitus, irritability, depression or fatigue.

While the other MTBI studies did not explicitly state that their samples had symptoms of PCS, all participants were reporting sufficient symptoms post MTBI to warrant a neuropsychological assessment. This suggests that all MTBI participants would have met the PCS inclusion critieria for Youngjohn, Burrows and Erdal's (1995) study.

Trueblood and Schmidt (1993) reported that 'most' of their sample had MTBI, however, they state that some participants may have been unconscious for slightly more than 30 minutes or had positive neuro-radiological findings. Similarly, Greve and Bianchini's (2006 set 'b') 'moderate to severe' TBI sample consisted of all TBI referrals that did not meet the MTBI criteria (i.e. were more severe).

3.1.2. Time since injury

Youngjohn, Burrows and Erdal (1995) required that their participants had experienced post brain injury symptoms for a minimum of six months prior to inclusion in the study. The remaining studies did not have a similar inclusion criterion except that participants were only assessed after any period of PTA. The mean period of time between injury and assessment across all five studies was 25.82 months.

3.1.3. External incentives to malinger

All identified participants had an external incentive in the form of substantial financial reward. Within each data set the exact nature of the incentive varied and

authors did not separate participants by type of monetary incentive. Incentives included third party compensation claims, worker's compensation and disability payments. None of the participants were recorded as being subject to criminal or competency to stand trial proceedings.

3.2. Methods of Identifying Definite Malingering

All of the reviewed studies utilised a forced-choice symptom validity test to determine whether or not their participants met the B1 (Slick, Sherman & Iverson, 1999) Criterion for definite malingering. Forced choice measures typically involve an exposure trial where participants are exposed to a set of pictures, words, or numbers and either asked to remember them or to judge whether or not they are pleasant. The participant is then either immediately or after a delay exposed to sets of two stimuli, one to which they will have already been exposed and asked to identify items previously presented. A person with severe memory impairments should, in theory, perform at least at chance level; therefore a score significantly below chance ($p < 0.05$) is suggestive of an effort to perform poorly and meets the B1 Criteria for definite malingering.

3.3. Methods of Identifying 'Probable' Malingering

With the exception of Binder's (1993) study all of the examined data sets provided percentages for 'probable' malingering. In order to meet the Slick, Sherman and Iverson (1999) criteria for 'probable' malingering an examinee must meet either two of the B criteria or one B criterion plus one C criterion (see table 1). All of the studies utilised the B2 criterion, where participants were required to score below standardised cut-offs for suspected effort on a well validated test of symptom validity (SVT). Scores below cut-offs on standardised SVTs are typically above chance

performance but lower than the floor performance of a severely brain injured population (e.g. Tombaugh, 1996). In addition to the B2 criterion, studies identifying 'probable' malingerers employed one of the following criteria: B3, B4, C2 or C5.

3.3.1. Criterion C5

Youngjohn, Burrows and Erdal (1995) and both of Greve and Bianchini's (2006) samples were examined according to the C5 (Slick, Sherman & Iverson, 1999) criterion. The C5 criterion dictates that there should be evidence of exaggerated or fabricated psychological dysfunction and accepts evidence in the form of validity scales embedded in measures of psychological adjustment. Both studies used the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Hathaway & McKinley, 1989), However, the two studies used the MMPI in different ways. Youngjohn, Burrows and Erdal (1995) used a profile analysis technique (e.g. Heaton, Smith, Lehman & Voght, 1978) where extreme levels of symptom reporting are viewed as suspicious. In contrast, Greve and Bianchini (2006) utilised the embedded 'Fake Bad Scale' (Lees-Haley, English & Glen, 1991).

3.3.2. Criterion C2

Trueblood and Schmidt's (1993) 'probable' malingerers met criterion C2 (see table 1) by displaying or reporting highly improbable symptoms which were discrepant with known patterns of brain functioning. These included symptoms such as zero grip strength and a full scale IQ of 90 despite 18 years in formal education.

3.3.3. Criteria B3/B4

Langeluddeck and Lucas' (2004) participants met either criterion B3 or B4, demonstrating a discrepancy between test data and either known patterns of functioning

or observed behaviour respectively. Participants met these criteria by showing an impairment on the WMS-III (Wechsler, 1997b) that was excessive either in the context of mild TBI or in the context of their level of functioning in their normal environment.

3.4. Study Design

Three of the studies (Binder, 1993; Langeluddeck & Lucas, 2004; Greve & Bianchini, 2006) identified malingerers for use in a 'known groups' design in order to validate another test of symptom validity. One study (Youngjohn, Burrows & Erdal, 1995) aimed to describe the nature of test results in a Post Concussion Syndrome sample in an attempt to better understand this client group. Trueblood and Schmidt (1993) set out to examine the characteristics of malingering and poor effort and its relationship with other test results during neuropsychological evaluation.

3.5. Methodological Critique

3.5.1. Biases in sampling

Of the five studies reviewed, only three (Langeluddeck & Lucas, 2004; Trueblood & Schmidt, 1993; Youngjohn, Burrows & Erdal, 1995) stated that they had sampled from consecutive referrals. Neither Greve and Bianchini (2006) nor Binder (1993) stated how they had selected their sample, raising the possibility that the sample had already been selected on the grounds of another criterion. While the primary aim of Greve and Bianchini's (2006) study was not to detect base rates of malingering within their sample they explicitly stated that they had identified them and related them to other base rate studies (e.g. Mittenberg, et al., 2010). This suggests that Greve and Bianchini (2006) were satisfied that they had not biased their sample by preselecting participants. Binder (1993) made no assertions regarding the base rates of malingering raising the possibility that his sample was pre-selected on some other grounds. Binder's

(1993) sample could have been pre-selected, for example, by only including subjects who had been administered a symptom validity test in a service where symptom validity tests are only administered in the presence of suspicious symptoms.

3.5.2. Choice of SVT and Alpha values

Both Binder (1993) and Youngjohn, Burrows & Erdal (1995) used the Portland Digit Recognition Test as a measure of symptom validity. The PDRT (Binder, 1993) has been found to have a sensitivity of .70 and specificity of .95 (Greve & Bianchini, 2006). Greve and Bianchini (2006) used the TOMM (Tombaugh, 1996) which has been shown to have a sensitivity of .65 and a specificity of .93 (Sollman & Berry, 2011). Langeluddeck and Lucas (2004) used the Warrington Recognition Memory Test (WRMT; Warrington, 1984); this has not been extensively validated as a measure of symptom validity, however, within student simulator studies (e.g. Iverson & Franzen, 1998), it has been found to have good levels of sensitivity (.90) and specificity (1). Trueblood and Schmidt (1993) used the Forced Choice Test (Hiscock & Hiscock, 1989) which has been found to have both sensitivity and specificity of .90 (Guilmette, Hart & Giulian 1993). Relatively low levels of test sensitivity suggest these tests may have under-diagnosed malingering within the selected studies.

Two studies (Binder, 1993; Youngjohn Burrows & Erdal, 1995) defined definite malingering as symptom validity test performance significantly below chance $p<0.1$, this contrasts with the alpha level provided by the other three studies where $p<0.05$ as suggested by Slick, Sherman & Iverson (1999). This contrast in acceptable alpha values indicates that Binder (1993) and Youngjohn, Burrows and Erdal (1995) may have over-estimated the base rate of definite malingering within their samples.

3.5.3. Referral sources

All of the reviewed studies indicate that participants were referred for neuropsychological evaluation following actual or suspected TBI. However, neither Greve and Bianchini (2006) nor Binder (1993) report on referral sources or referral questions. Similarly, Trueblood and Schmidt (1993) and Youngjohn, Burrows and Erdal (1995) state that referrals came from a variety of sources, including other medical professionals, insurance companies and attorneys with no qualification of how many participants came from which source. Without details of the referral sources and number of participants resulting from each source it is difficult to assess the extent to which the acquired US sample is reflective of the US compensation seeking TBI population.

3.5.4. Referral questions

Trueblood and Schmidt (1993) state that none of their participants were referred specifically regarding queries of malingering; raising the issue that the other authors may have received referrals because malingering was suspected. If suspected malingering was the referral question for any of the other participants it would inflate the base rate figures for this review. Furthermore, all of the neuropsychological assessment services collecting data were private practices, suggesting that referrers may have had a choice as to where they referred their clients. If a service had a reputation for 'detecting malingering' there is a high potential for referrals, and therefore any resulting sample, to be biased. The only study to state the referral question associated with participants was that of Langeluddeck and Lucas' (2004) Australian study where all participants were referred in relation to claims for compensation. While all participants across the five studies were seeking compensation or benefits, it is not clear whether their assessment was directly related to their claim. If Langeluddeck and

Lucas' (2004) study was the only one where participants were aware that their assessment would have a direct bearing on their claim for monetary rewards this could inflate the base rate of malingering within their sample relative to the other samples.

3.5.5. Cross-cultural differences in litigation

Four of the five selected studies were based in North America, with the remaining study (Langeluddeck & Lucas, 2004) based in Australia. The United States of America is widely regarded as the most litigious society in the world, with the costs associated with direct tort in 2003 amounting to 2.24% of GDP (Tillinghast, 2003). In contrast, the direct tort costs for the UK in the same year were 0.7% of GDP (Tillinghast, 2003), suggesting a significant difference in litigation contexts between the two countries. As a result of these differences it is unlikely that the base rates of malingering are comparable between the two countries and this review should be considered as an indicator of US base rates only.

Litigation contexts differ between states within the US, the samples selected for this review were from Oregon (Binder, 1993), California (Greve & Binachini, 2006), Colorado (Trueblood & Schmidt, 1993) and Arizona (Youngjohn, Burrows & Erdal, 1995). The US Tort-Liability Index (2010) ranked states in terms of tort costs and tort litigation risks, where a score of 50 (New Jersey) represented the highest cost and risk and 1 the lowest (Alaska). In terms of the reviewed samples California scored particularly high for cost and risk, ranked 41 out of 50. Arizona was the lowest risk, ranking 16 and Colorado and Oregon were somewhere in-between ranking 32 and 34 respectively. Higher financial rewards are associated with an increased risk of effort test failure, particularly in Mild TBI samples (Bianchini, Curtis & Greve, 2006). Therefore the present review would be more accurately applied to the states with higher litigation

risk and costs as it may overestimate malingering base rates if applied to lower risk states.

Given that US litigation rates are the highest in the world, it seems unlikely that base rates of malingering in the US can be compared with other countries such as Australia. A survey of Neuropsychologists estimated malingering base rates within Australian mild head injury cases to be just 23% (Sullivan, Lange, & Dawes, 2005). This is significantly lower than the figure of 39% for North American mild head injury cases as presented by Mittenberg, Patton, Canyock and Condit (2002) using the same survey. However, the homogeneity of Langeluddeck and Lucas' (2004) sample suggests it is likely to be representative of Australian, litigating Mild Head Injury examinee's referred for assessment in relation to a compensation claim.

3.6. Results

The identified percentages of participants 'definitely' and 'probably' malingering within each data set are presented in Table 2. Across the six data sets (N= 602) 6.96% were identified as 'definite malingerers'. Between samples, the proportion of definite malingerers ranged from 1.38% (Greve & Bianchini, 2006, set 'a') to 17.47% (Binder, 1993).

Across the five data sets (N=499) identifying 'probable malingerers' 18.44% fit the criteria. Between samples the proportion of 'probable' malingerers ranged from 5.66% (Trueblood & Schmidt, 1993) to 24.83% (Greve & Bianchini, 2006 set 'a').

In order to take account of cross cultural differences in litigation data from the Australian study (Langeluddeck & Lucas, 2004) were removed from the descriptive analysis. Unexpectedly the removal of Australian data made little difference to the

overall average numbers of those classified as definite or probable malingerers. Removing Australian data resulted in a slight increase (0.57%) in the proportion of the sample diagnosed as definite malingerers and, conversely, a slight decrease (1.34%) in the proportion of the sample diagnosed as probable malingerers. As noted above, Langeluddeck and Lucas' (2004) participants were all assessed in direct relation to an active compensation claim, suggesting that their base rate may have been inflated in comparison to other samples whose assessment may have not been directly related to their compensation claim.

3.7. Conclusion

Data from 503 litigating TBI North Americans revealed an average classification base rate of 7.55% for definite and 17% for probable malingering. While these figures may be depressed by poor test sensitivity, they are also likely to be slightly inflated due to the use of high alpha values ($p<0.1$) and the litigating culture of the geographical regions covered by the sample. The results from one Australian study suggested that Australian base rates of malingering may be similar to those of the US. However, further research is needed to clarify this relationship as the Australian sample's figures were also likely to be inflated by the characteristics of their sample.

A lack of any identified research into UK malingering base rates suggests that research in this area is much needed. It is hypothesised that UK base rates are likely to be much lower than those of the US due to lower incidence and smaller financial incentives associated with litigation.

Table 2. The Base Rates of Probable and Definite Malingering

#	Authors	Date	Location	Type of ABI[1] Mean Months Since Injury (SD)	Sample size (% Female) Mean Age	Forced Choice Test[2]	Definite N (%) (p<)	Probable criteria[3]	Probable
1	Youngjohn, Burrows & Erdal	1995	North America	PPCS 32.7 (38.3)	54 (41%) 39	PDRT	8 (15%) (p<0.1)	B2 C5	10 (18%)
2a	Greve & Bianchini	2006	North America	Mild TBI: 22.2 (24.4)	145 (32%) 40	TOMM	2 (1.38%) (p<0.05)	B2 C5	36 (24.83%)
2b				Mod-Sev TBI 22.2 (24.4)	95(19%) 35	TOMM	2 (2.11%) (p<0.05)	B2 C5	16 (16.84%)
3	Langeluddeck & Lucas	2004	Australia	Mild TBI 39 (not reported)	99 (29%) 35	WRMT	4 (3.96%) (p<0.05)	B2 B4/ B3	24 (23.76%)
4	Binder	1993	North America	Mild TBI 24 (21.4)	103 (32%) 40	PDRT	18 (17.47%) (p<0.05)	N/A	N/A

[1] PPCS: Persistant Post-Concussion Syndrome, TBI: Traumatic Brain Injury.
[2] PDRT: Portland Digit Recognition Test (Binder, 1993), TOMM: Test of Memory Malingering (Tombaugh , 1996), WRMT: Warrington Recognition Memory Test (Warrington, 1984), FCT: Forced Choice Test (Hiscock & Hiscock 1989)
[3] Slick, Sherman and Iverson (1999) criteria for 'probable malingering' (see table 1)

5	Trueblood & Schmidt	1993	North America	Mild-Mod TBI 14.8 (not reported)	106 (64%) 34	FCT	8 (7.55%) ($p<0.05$)	B2 C2	6 (5.66%)
	Total	Excl	Australia	Mean months since injury=23.18	503 (37.6) 37.6		38 (7.55%)		68/400 (17%)
	Total	Inc	Australia	Mean months since injury= 25.82	602 (36%) 37		42 (6.98%)		92/499 (18.44%)

4. Discussion

A systematic search of the literature revealed five independent studies which had applied the Slick, Sherman and Iverson (1999) criteria to determine malingering base rates within Traumatic Brain Injury (TBI) samples. Of the five studies, four sampled from North American, litigating TBI populations where base rates of 'definite' malingering were, on average, 7.55% ranging from 1.38% (Greve & Bianchini, 2006) to 17% (Binder, 1993). Within the North American samples average base rates of 'probable' malingering were 17%, ranging from 5.66% (Trueblood & Schmidt, 1993) to 33% (Youngjohn, Burrows & Erdal, 1995). Due to methodological issues, including inflated alpha levels and the potential for referral bias, these figures are likely to be an overestimation of the malingering base rates in North America. The high frequency and value of awards in North America (US Tort-Liability Index, 2010) make the incentives to malinger higher than anywhere else in the world. Bianchini, Curtis and Greve (2006) established a 'dose-response relationship' between the monetary value of an incentive and the likelihood of symptom validity test failure whereby as incentives increase in monetary terms the likely hood of symptom validity test failure increases. Therefore, it is likely that the base rates of 'definite' and 'probable' malingering in other countries are significantly lower.

4.1. Scope of findings

The present review aimed to sample base rate data across countries: unfortunately, only one non-US study meeting the inclusion criteria was found. Langeluddeck and Lucas (2004) diagnosed 'definite' in 3.96% and 'probable' malingering in 23.76% of their Australian, litigating, mild TBI sample. Due to the cross-cultural differences in litigation Langeluddeck and Lucas' sample were excluded from the average base rate statistics provided above. However, unexpectedly the

Australian sample's base rates fell within the range of base rates provided by the North American samples. This may be because all of Langeluddeck and Lucas' participants were being assessed specifically in relation to their compensation claim. In order to make a comparison between Australian and North American base rates of malingering more Australian research is needed. A lack of UK base rate studies have resulted in an over-reliance on North American base rate data (e.g. BPS, 2009) in both estimating malingering base rates and interpreting symptom validity test results. Therefore UK research into base rates of definite and probable malingering is essential if the BPS wish to continue to assert that Symptom Validity Tests should be used in routine clinical practice (BPS, 2009).

4.2. Clinical Implications

The North American base rates of definite (7.55%) and probable (17%) as found in the present review are significantly lower than previous estimates have suggested. Even if it is assumed that all participants classified as 'probable' malingerers are malingering the base rate figure would be 24.55%. This figure is significantly lower than Larabee's (2003) highly cited 40% and Mittenberg et al.'s (2002) 39% base rates for malingering in litigating mild TBI samples. Whilst the reviewed studies employed symptom validity tests with limited sensitivity to malingering, both Larabee (2003) and Mittenberg et al.'s (2002) base rate estimates would have been equally affected by this issue, suggesting that the base rate of malingering remains lower than previously estimated.

When the base rate figure of 24.55% is applied to commonly used symptom validity tests a problem arises. The most commonly used symptom validity test in the UK (McCarter, et al., 2009) is the Test of Memory Malingering (TOMM; Tombaugh,

1996), followed by the Word Memory Test (WMT; Green, 2005). According to a recent meta analysis by Sollman and Berry (2011) the sensitivity and specificity of the TOMM (1996) are 65.4% and 93.8% respectively. Sollman and Berry (2011) found the sensitivity and specificity of the WMT (Green, 2005) to be 75.1 and 69.4 respectively. With a base rate of 24.55% the TOMM's Positive Predictive Accuracy (PPA) falls to 77% meaning that 23% of examinees would be wrongly classified as malingerers. When the base rate of 24.55% is applied to the WMT (Green, 2005) its PPA drops to 44%, meaning that someone classified as malingering by the WMT is more likely to have been inaccurately classified than actually malingering. It should be noted that due to methodological concerns, Sollman and Berry (2011) excluded several published studies (e.g. Green, 2007), which estimated the WMT's sensitivity and specificity rates to be higher.

The Positive Predictive Accuracy for all symptom validity tests will vary as a function of the base rate within the population being assessed. Therefore, if as hypothesised, UK base rates for malingering are lower than that of the North American litigating TBI population sampled here, the PPA is likely to be significantly lower in UK populations. The BPS (2009) recommend that symptom validity tests should be administered as routine to all persons presenting for neurocognitive assessment unless they have a severe sensory or motor impairment, severe dementia or are experiencing post traumatic amnesia. Populations with low or no external incentive to perform poorly are likely to have lower base rates of malingering which would further decrease the positive predictive accuracy of symptom validity tests. In cases where there was no external incentive to perform poorly, malingering would not be diagnosable according to the Slick, Sherman and Iverson (1999) criteria. However, clinicians could attribute

test failure to invalid responding or insufficient effort where clients are in fact employing their 'best effort'.

A multi-test approach to symptom validity testing has previously been recommended to increase diagnostic accuracy in assessment (e.g. BPS, 2009; Slick, Sherman & Iverson, 1999). While it may seem appropriate to assume that a combination of tests would increase positive predictive accuracy, tests of symptom validity, particularly those employing a forced-choice paradigm are likely to be highly correlated (Rosenfeld, Sands & Van Gorp, 2000). Therefore, the false positives on one test may simply be repeated by the other test. Further research is needed to examine the relationship between different types of symptom validity tests.

While tests of symptom validity are widely regarded as very easy, they are still tests of recognition memory requiring the ability to attend to and store presented information for a period of time (BPS, 2009). It follows that persons with very severe attentional or memory impairments may fail these tests (Lezak, 2004) despite full effort. Furthermore, if effort is considered as a vector with both a magnitude and direction (BPS, 2009), symptom validity test failure may not represent an effort to mislead, but, instead, no effort to succeed. Cripe (2002 cited in Lezak, 2004) suggests a number of reasons someone might 'fail' a symptom validity test, including pain, fatigue, frustration with the testing process or examiner-examinee relationship. Such a lack of motivation or attentional capacity may result in the respondent failing to attend to the test material and result in random responding. If participants were responding at random on recognition trials, chance would suggest that a significant number would score below 'floor' cut offs and 5% might score significantly below chance levels.

4.3. Conclusion

The reviewed studies suggest that in North American litigating TBI samples base rates of individuals meeting the slick criteria for either definite or probable malingering is no higher than 24.55%. This is likely to be an overestimation of North American base rates and should not be generalised to outside of North America, except that base rates in other countries are likely to be no higher than 24.55%. This base rate figure is significantly lower than previous estimates suggesting that the positive predictive accuracy of symptom validity tests may be severely compromised.

4.4. Research Implications

In order for symptom validity tests to be reliable further research is needed to ascertain the likely base rates of malingering outside of North America and within samples with different levels of external incentive. Further development of symptom validity tests could include alteration of cut-off scores to account for a lower base rate, increasing specificity and therefore positive predictive accuracy. While questions regarding the construct validity of symptom validity tests were outside of the scope of the present review, further research in this area is called for. Such research could include an examination of the factors associated with symptom validity test failure in samples with and without external incentives and examine the relationship between different symptom validity tests. An understanding of the relationships between symptom validity tests and the development of uncorrelated measures could be used to increase the positive predictive accuracy of diagnosis. Further exploration of the construct validity of 'malingering' and its relationship with 'effort' may enhance understanding of what symptom validity test failure actually means.

5. References

* Reviewed within the present paper

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar?—a meta analysis of individual differences in detecting deception. *Forensic Exam, 15*, 6 –11.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders. Text Revision* (4th Ed.). Washington, D.C: American Psychiatric Association.

Ardolf, B. R. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist , 21*, 899-916.

Babikian, T., & Boone, K.B. (2007). Intelligence tests as measures of effort. In K. Boone (Ed.), *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective.*(pp 103-127). New York: Guilford Press.

Bianchini, K. J., Curtis, K. L., & Greve, K. W. (2006). Compensation and malingering in traumatic brain injury: A dose-response relationship. *The Clinical Neuropsychologist, 20*, 831-847.

*Binder, L. M. (1993). Assessment of malingering after mild head trauma with the portland digit recognition test. *Journal of Clinical and Experimental Neuropsychology, 15*,170-182.

Binder, L. M., & Kelly, M. P. (1996). Portland digit recognition test performance by brain dysfunction patients without financial incentives. *Assessment, 3*, (4), 403-409.

British Psychological Society; Professional Practice Board. (2009). *Assessment of Effort in Clinical Testing of Cognitive Functioning for Adults.* Leicester, UK: British Psychological Society.

Curtis, K. L., Thompson, L. K., Greve, K. W., & Bianchini, K. J. (2008). Verbal fluency indicators of malingering in traumatic brain injury: Classification accuracy in known groups. *The Clinical Neuropsychologist, 22*, 930-945.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgement. *Science, 243,* 1668-1674.

Drob, S. L., Meehan, K. B., & Waxman, S. E. (2009). Clinical and conceptual problems in the attribution of malingering in forensic evaluations. *The Journal of the American Academy of Psychiatry and the Law, 37,* 98-106.

Green, P. (2004). *Medical Symptom Validity Test for Windows: User Manual and Program,* Edmonton, Canada: Green's Publishing Inc.

Green, P. (2005). *Green's Word Memory Test for Windows; User's Manual- Revised.* Edmonton, Canada: Green's Publishing Inc.

Green, P. (2007). Spoiled for choice: Making comparisons between forced-choice effort tests. In K. Boone (Ed.), *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective.* (pp 50-77). New York: Guilford Press.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6,* (3), 218-224.

*Greve, K. W., & Bianchini, K. J. (2006). Classification accuracy of the portland digit recognition test in traumatic brain injury. *The Clinical Neuropsychologist, 20,* 816-830.

Greve, K. W., Bianchini, K. J., Love, J. M., Brennan, A., & Heinly, M. T. (2006). Sensitivity and specificity of MMPI-2 validity scales and indicators to malingered

neurocognitive dysfunction in traumatic brain injury. *The Clinical Neuropsychologist, 20,* 491-512.

Grote, C. L., Kooker, E. K., Garron, D. C., Nyenhuis, D. L., Smith, C. A., & Mattingly, M. L. (2000). Performance of compensation seeking and non-compensation seeking samples on the victoria symptom validity test: Cross-validation and extension of a standardization study. *Journal of Clinical and Experimental Neuropsychology*, *22*, (6), 709-719.

Hathaway, S. R., & Mckinley, J. C. (1989). *The Minnesota Multiphasic Personality Inventory-2: Manual for Administration and Scoring.* Minneapolis, MS: University of Minnesota Press.

Heaton, R. K., Smith, H. H., Lehman, R. A. W., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46,* (5), 892-900.

Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 11,* 967-974.

Inman, T. H., Vickery, C. D., Berry, D. T. R., Lamb, D. G., Edwards, C. L., & Smith, G.T. (1998). Development and initial validation of a new procedure for evaluating adequacy of effort given during neuropsychological testing: The letter memory test. *Psychological Assessment, 10,* 128-149

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A fake bad scale on the MMPI-2 for personal injury claimants. *Psychological Reports, 68,* 203-210.

*Langeluddecke, P. M., & Lucas, S. K. (2004). Validation of the rarely missed index (RMI) in detecting memory malingering in mild head injury litigants. *Journal of Forensic Neuropsychology, 4*, (1), 49-64.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist , 17*, (3), 410-425.

Lezak, M. D. (2004). *Neuropsychological Assessment*. (4th Ed.). New York: Oxford University Press.

Lipman, F. D. (1962). Malingering in personal injury cases. *Temple Law Quarterly, 35*, 41-162

Meyers, J. E., & Volbrecht, M. (1998). Validation of reliable digits for detection of malingering. *Assessment, 5*, (3), 303-307.

McCarter, R. J., Walton, N. H., Brooks, D. N., & Powell, G. E. (2009). Effort testing in contemporary UK neuropsychological practice. *The Clinical Neuropsychologist, 23*, 1050-1066.

McQuillan, L. J. & Abramyan, H. (2010). *U. S. Tort Liability Index: 2010 Report*. San Francisco: Pacific Research Institute.

Mild Traumatic Brain Injury Committee of the Head Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation Medicine. (1993). Definition of mild traumatic brain injury. *Journal of Head Trauma and Rehabilitation, 8*, 86-87.

Millis, S. R. (1992). The recognition memory test in the detection of malingered and exaggerated memory deficits. *The Clinical Neuropsychologist*, 6, (4), 406-414.

Millis, S. R., Putnam, S. H., Adams, K. M., & Ricker, J. H. (1995). The california verbal learning test in the detection of incomplete effort in neuropsychological evaluation. *Psychological Assessment 7,* (4), 463-471.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24,*(8), 1094-1102.

Ord, J. S., Greve, K. W., & Bianchini, K. J. (2008). Using the wechsler memory scale-III to detect malingering in mild traumatic brain injury. *The Clinical Neuropsychologist, 22,* 689-704

Rogers, R. (1997). Introduction. In R. Rogers (Ed.) *Clinical Assessment of Malingering and Deception* (2nd ed., pp. 1-19). New York: Guilford Press.

Rohling, M. L., Green, P., Allen, L. M., & Iverson, G. L. (2002). Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. *Archives of Clinical Neuropsychology*, *17*, 205-222.

Rosenfeld, B., Sands, S. A., & Van Gorp, W.G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology, 15,* (4) 349-359.

Slick, D., Strauss, E., & Spellacy, F. (1996). Victoria symptom validity test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology, 18,* 911-922.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13,* 545-561.

Sollman, M. J., & Berry, D. T. R. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology, 26,* 774-789.

Tillinghast. (2003). *U. S. Tort Costs: 2003 Update: Trends and Findings on the Costs of the U. S. Tort System.* U.S: Tillihast-Towers Perrin.

Tombaugh, T. N. (1996). *Test of Memory; TOMM.* North Tonawanda, NY: Multi-Health Systems Inc.

*Trueblood, W. & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology, 15,*(4), 578-590.

Van Gorp, W.G., Humphrey, L. A., Kalechstein, A. L., Brumm, V. L., McMullen, W. J., & Stoddard, M. A. (1999). How well do standard clinical neuropsychological tests indentify malingering? A preliminary analysis. *Journal of Clinical and Experimental Neuropsychology, 21,* 245-250.

Warrington, E. K. (1984). *Recognition Memory Test*. Berkshire, UK: NFER-Nelson.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale-III.* San Antonio: The Psychological Corporation.

Wechsler, D. (1997b) *Wechsler Memory Scale 3rd Edition Manual*. San Antonio: The Psychological Corporation.

Wilson, B. A., Alderman, N., & Burgess, P.W. (1996). *Behavioural Assessment of the Dysexecutive Syndrome.* Bury St. Edmunds, UK: Thames Valley Test.

*Youngjohn, J. R., Burrows, L., & Erdal, K. (1995). Brain damage or compensation neurosis? The controversial post-concussion syndrome. *The Clinical Neuropsychologist*, 9, (2), 112-123.

Research Paper

Effort Test Results: The Effect of Informed Consent in a Clinical Sample

Effort Test Results: The Effect of Informed Consent in a Clinical Sample

Alice Nicholls

Abstract

Introduction: The British Psychological Society (BPS, 2009) advises that all patients presenting for neuropsychological assessment should be informed that they will be tested for effort. Effort or Symptom Validity Tests (SVTs) are constructed and normed using uninformed samples, therefore, the act of informing patients may invalidate their assessment. The present study aimed to determine whether informing patients they will be assessed for effort affects their effort test results. Previous research had investigated the effect of warning student simulators they would be assessed for effort, finding that warned simulators outperformed un-warned simulators.

Method: Patients presenting for neuropsychological assessment at four UK neuropsychology services were invited to participate. Participants were randomly allocated to either 'informed' or 'uninformed' conditions and assessed as appropriate to their clinical need using tests including: the Test of Memory Malingering (Tombaugh, 1996); tests of memory and other embedded measures of effort.

Results: Practical difficulties associated with recruitment led to an inadequate sample size for statistical power. The null hypothesis was neither rejected nor retained and data collection is ongoing.

Discussion: Recruitment difficulties highlighted discrepancies in the attitudes towards and use of SVTs within UK clinical practice. It also identified a lack of consensus amongst clinicians in the balancing of BPS best practice guidelines against ensuring test validity. Implications for clinical practice and further research are discussed.

1. Introduction

1.1. Neuropsychological Assessment

Within the UK, Neuropsychological assessments may be carried out within the National Health Service (NHS) after an acquired brain injury or during the course of a congenital or degenerative condition to aid diagnosis, gauge the extent of functional impairment and identify treatment, rehabilitation and future care needs. While a client may use a resulting report to support a claim for disability benefits or compensation, this would not be the sole purpose of the examination.

Clients wishing to evidence a claim for litigation through neuropsychological assessment are typically referred to private neuropsychological assessment services by their insurance companies. However, the information gathered may also be used to identify further appropriate treatment for the individual.

1.2. Symptom Validity

In order for neuropsychological test results to be reliable and valid it is necessary for the client to employ their best effort. As such, most neuropsychological test manuals instruct the examinee to try their best throughout testing (e.g. Wechsler, 2010). Failure to deploy their best effort or making an effort to deceive the examiner would result in inaccurate test results. US and UK studies have found that 53% of the variance in neuropsychological test results can be explained by symptom validity test results (Green, Rohling, Lees-Haley & Allen, 2001 and Moss, Jones, Folkias & Quinn, 2003 respectively). In order to control for this potential bias in test results much research (e.g. Boone, 2007) has examined ways of measuring symptom validity.

While there is no base rate statistic available for malingering in the UK the British Psychological Society (BPS; 2009) asserts that where there is financial incentive to deceive, this is likely to be high. As symptom exaggeration or invalidity can arise in a variety of contexts the BPS (2009) advises that symptom validity testing should be standard clinical practice in all neuropsychological assessments.

1.3. Symptom Validity Tests (SVTs)

SVTs usually employ forced choice paradigms, basing judgements of symptom validity on the statistical probability of getting a certain proportion of responses right by chance. When an individual makes fewer correct responses than chance would dictate the validity of the client's symptoms is called into question. Pankratz (1988) suggests that the subject of the forced choice test should be determined by the presentation of the client. For example: in the case of memory difficulties the test should appear to be a test of memory as this motivates the malingering client to demonstrate their reported deficiency. SVTs such as the Test of Memory Malingering (TOMM; Tombaugh, 1996) employ a forced choice recognition paradigm for complaints of poor memory, are in widespread use in the UK (McCarter, Walter, Brooks & Powell, 2009) and are endorsed by the British Psychological Society (BPS; 2009).

1.4. The Confusion between 'Insufficient Effort' and 'Malingering'

Throughout much of the literature (e.g. BPS, 2009) the terms 'insufficient/poor effort' and 'malingering' are used interchangeably. The failure to differentiate between effort and malingering seems to have arisen from the unwillingness of psychologists to use the value laden, and potentially damaging, term 'malingering'. Due to the unattractiveness of the term it has become common practice for 'poor effort' 'suspect effort' or 'insufficient effort' to be used as euphemisms or synonyms for 'malingering'.

These euphemisms are inherently misleading because, in order to successfully malinger, clients must deploy good levels of effort. Furthermore, tests of malingering or symptom validity tests were not designed to measure the levels of effort deployed by examinees, rather, they were designed specifically to assess whether or not the client was malingering. As such, symptom validity tests should not be referred to as measures of effort, nor should they be expected to assess effort (Slick, Sherman & Iverson, 1999).

1.5. The Issue of Informed Consent

The BPS (2009) advises that clients should be informed that the level of effort they deploy during assessment will be measured. However, all symptom validity tests (SVTs) prohibit informing the client that they will be tested for effort. Furthermore, most SVTs are presented in a way that actively encourages deception. For example the TOMM (Tombaugh, 1996) administration manual dictates that the assessor should describe the test as a measure of the client's ability to learn and remember pictures of common objects.

The BPS (2009) acknowledges that the validity of SVT results would be questionable if the test was identified to the client. They advise that the client is informed they will be tested for effort at the beginning of a battery of neuropsychological tests and not directly before SVTs. However, tests of symptom validity have been standardised using samples who were not warned or informed that the validity of their responses were being examined, therefore the validity of the tests administered under informed conditions is questionable (Boone, 2007).

A recent survey of members of the BPS found that while 73% of respondents working in medico-legal settings used SVTs for at least half of their assessments; only 16% of respondents working solely in clinical settings used SVTs to this extent

(McCarter, Walter, Brooks & Powell, 2009). McCarter et al. (2009) did not survey respondents on whether they inform clients that they will be tested for effort. However, a US survey of National Academy of Neuropsychology members (Sharland & Gfeller, 2007) found that only 27% of respondents 'always or often' provide a warning that effort will be tested.

There are several reasons why it may be good practice to inform clients that they will be tested for effort or symptom validity: a warning may decrease symptom fabrication or exaggeration which would make cognitive test results more representative of the patient's true level of functioning. It may also be considered unethical to proceed with tests for which the client has not given their informed consent. In contrast there are concerns that patients who are malingering will apply a more sophisticated approach to their test responses by *"suppressing the tendency to do devastatingly poorly on measures they perceive to be easy"* (Gunstad & Suhr, 2001, p.402) or by attempting to identify which test is a test of effort (Boone, 2007).

1.6. Previous Research

No previous research has explicitly examined the effect of informing clients that they will be tested for effort on their symptom validity test results. However, several studies have examined the impact of coaching or warning simulating participants. Studies that examine coaching typically give participants guidance or clues as to how they can perform well on symptom validity tests while studies examined the effect of a warning tend to give participants less information but warn them that poor effort or 'faking' may be detected by tests.

1.6.1. Coaching studies

Studies examining the effect of 'coaching' seem to have been provoked by increasing concerns that attorneys may have coached litigants to pass symptom validity tests (Wetter & Corrigan, 1995). Most test coaching studies, using simulated malingering designs have found that coached participants perform better than 'naive' malingerers but worse than full effort controls (e.g. Dunn, Shear, Howe & Ris, 2003; Rose, Hall, Szalda-Petree & Bach, 1998; Martin, Bolter, Todd, Gouvier & Niccolls, 1993; Colman, Rapport, Millis, Ricker & Farchoine, 1998). Some authors concluded that the sensitivity of their symptom validity test was compromised (e.g. Colman et al., 1998) while others noted that while 'coached' malingerers outperformed naive malingerers, it was still possible to differentiate between coached malingerers and full effort controls (e.g. Powell, Gfeller, Oliveri, Stanton & Hendricks, 2004).

1.6.2. Warning studies

BPS (2009) advice on effort testing suggests that participants should be told that they will be tested for effort without providing clues as to how they might identify or pass symptom validity tests. A database search for studies examining the effect of warning or informing participants that they would be tested for effort revealed 10 articles. None of the identified studies examined the effect of warning clinical samples that they would be tested for effort. All studies utilised analogue designs where (non-UK) undergraduate students were either asked to act as if they had experienced a traumatic brain injury and were faking the extent of their injuries (malingering group) or to complete tests using their best effort (control group). A proportion of the 'malingering' groups were given warnings regarding the potential detection of malingering.

The content of the warning varied between studies. Four studies (Suhr, 2002; Suhr & Gunstad, 2000; Jelicic, Merckelbach, Candel & Geraerts, 2007; Jelicic, Ceunen, Peters & Merckelbach, 2011) used Suhr and Gunstad's (2000) instructions, where a brief vignette about how the participant acquired their head injury, why they decided to exaggerate their symptoms and symptoms of brain injury are detailed. Participants in the "warned" condition were also given the following warning:

"At least one of the tests you will be given is designed to catch you faking, because it's easier than it looks. Be careful." (Suhr & Gunstad, 2000 pp 424)

Similarly, three studies (Johnson, Lesniak-Karpiak, 1997; Sulivan, Keane & Deffenti, 2001; Sulivan, Deffenti & Keane, 2002) warned participants that the tests they were taking may detect any exaggeration or faking. These studies did not provide any additional information on how the test looked nor how it might detect faking.

Three further studies (Schenk & Sulivan, 2010; Sulivan & Richer, 2002; King & Sulivan, 2009) warned simulating participants that the tests they were taking may be able to detect malingering and that if malingering was detected there would be a punishment. The punishments varied between studies, with Sulivan and Richer's (2002) participants told to imagine that detection of malingering would lead to criminal prosecution. Schenk and Sulivan's (2010) and King and Sulivan's (2009) participants were told they would be punished personally either by the removal of course credit (King & Sulivan, 2009) or by being removed from the prize draw and recorded as a *"poor psychology participant"* (Schenk & Sulivan, 2010, p. 754) on the course database.

The information given to participants receiving Suhr's (2000) warning, goes beyond informing the participant that they will be tested for effort as it provides a clue

as to how to avoid detection. This would be excessive in a clinical context where the aim of the clinician would be to inform clients that they would be tested for effort without compromising test security or validity. Similarly the warnings provided by the three studies informing participants there will be a punishment (Schenk & Sulivan, 2010; Sulivan & Richer, 2002; King & Sulivan, 2009) are not representative of the instructions suggested by the BPS (2009). In samples of litigating brain injury patients it is likely that both an incentive to exaggerate or fake symptoms (e.g. compensation) as well as a deterrent (e.g. prosecution for fraud, loss of employment) are present for each individual, varying as a function of individual circumstances. However, it seems that in Schenk and Sullivan's (2010) study the perceived costs of being caught malingering (being publicly labelled as a poor participant, loosing course credit and being excluded from a prize draw) may have outweighed the perceived potential gains of successful malingering (being entered into a prize draw to win $50 book vouchers). Indeed, 82% of participants receiving the 'high risk' warning admitted to deciding not to 'malinger' and instead performed at full effort. This suggests that Schenk and Sullivan's (2010) results may be more symptomatic of the particular costs and benefits presented to their participants than the effect of being informed that they would be tested for effort alone.

The three studies (Sulivan, Keane & Deffenti, 2001; Sulivan, Deffenti & Keane 2002; Johnson, Lesniak-Karpiak, 1997) which provided a simple warning that exaggeration or faking may be detected by tests provide closer approximations of 'informing' rather than 'coaching'. However, the use of simulators who have already been asked to 'fake' head injury symptoms is in itself problematic (Larabee, 2007): more so when they are then warned that malingering may be detected. Researchers attempted to improve ecological validity by offering participants a 'reward' for convincing malingering (Sulivan, Keane & Deffenti, 2001; Sulivan, Deffenti & Keane,

2002), checking post-hoc that participants understood the instructions and complied with them and excluding participants who admitted to misunderstanding or failing to comply from analysis (Sulivan, Keane & Deffenti, 2001; Sulivan, Deffenti & Keane, 2002; Johnson, Lesniak-Karpiak, 1997). Of these three studies, two (Sulivan, Keane & Deffenti, 2001; Sullivan, Deffenti & Keane, 2002) found no difference in psychometric test performance between naive and warned malingerers with both malingering groups performing significantly worse than 'full effort' controls. In contrast, Johnson and Lesniak-Karpiak (1997) found that warned group test results frequently approximated controls with unwarned malingerers performing significantly worse. Johnson and Lesniak-Karpiak conclude that their warning was effective in reducing malingering behaviour within their sample. Several factors may have contributed to the difference in results of the three studies: Johnson and Lesniak-Karpiak included a greater number of participants; did not provide an incentive to malinger; and used a North American sample while the other two studies sampled from an Australian student population. The American study used motor tasks as a measure of malingering while the two Australian studies examined results on a memory test.

1.6.3. Interpretation of the 'warning effect'

Within the literature there is some disagreement regarding the possible interpretation of any detected 'warning' effect. Youngjohn, Lees-Haley and Binder (1999) reviewed a number of studies where simulating participants were warned that they would be tested for effort; they found that the warned groups consistently scored better on tests than naive malingerers. They argued that warned participants were able to 'fake' more convincingly and therefore, examinees should not be warned they will be tested for effort. In contrast, Schenk and Sullivan (2010) suggested the reason participants in 'warned' groups score better than those in naive groups is that they have

been successfully deterred from malingering. In their own study they found most warned simulators decided not to malinger and performed similarly to full-effort controls. However, Schenk and Sullivan (2010) found that the few (19%) participants who received a high level warning and decided to malinger performed better than the 'naive' malingerers. This suggests that, for a small minority of examinees who are not sufficiently deterred from malingering, sensitivity of symptom validity tests may be compromised by a warning.

1.7. Summary

In summary, limited previous research into the effect of informing student simulators that they will be tested for effort has produced mixed results. Furthermore, there is disagreement regarding the interpretation of such results. A lack of research into the effect of informing UK clinical samples that they will be tested for effort makes it impossible to assess whether following the BPS (2009) guidance would result in reducing the sensitivity of symptom validity tests or decreasing the rate of invalid test results.

1.8. Primary Research Aims

The present research aims to determine whether informing UK clients presenting for neuropsychological assessment that they will be tested for effort affects their performance on a test of symptom validity, tests of memory and embedded measures of effort. Furthermore, if an effect is detected, the present study will investigate whether this effect is moderated by litigation status, extent of brain injury, state and trait anxiety and trait depression. In order to gain a sample that is representative of clinical cases within the UK both litigating and non litigating clients were invited to participate. Non-litigating neuropsychology clients are often excluded from studies investigating

symptom validity test results on the basis that they are assumed to have no external incentive to malinger and therefore should score about cut-offs on symptom validity tests. However, the BPS (2009) suggests that many non-litigants may have substantial external incentive to malinger (i.e. disability, pension or incapacity benefits) and may therefore meet the first of the necessary criteria for malingering (Slick, Sherman & Iverson, 1999). Furthermore, previously reported non-litigating samples have shown significant variability in symptom validity test performance. For example, in a predominantly non-litigating sample reported by Green (2007) 33% failed at least one of two symptom validity tests.

1.9. Secondary Research Aims

While the BPS (2009) advises that examinees should be informed they will be assessed for effort they acknowledge concerns that this may cause the client anxiety. Therefore the present research will examine the effect of informing clients that they will be tested for effort on their self reported state anxiety levels.

For the purposes of the main research question the Benton Visual Retention Test (BVRT; Sivan, 1992) will be used as a measure of visual memory. However, previous research suggests examinees simulating a brain injury make more distortion errors than those with genuine head injuries who make mainly omission errors (Benton & Spreen, 1961). Therefore the present study will examine whether results on the BVRT can be used to differentiate between clients with and without external incentives to perform poorly. Many symptom validity tests employ a forced choice paradigm, which, with information either from attorneys or personal research are easy for dishonest examinees to identify (Nitch & Glassmire, 2007). If the BVRT was validated as a test of symptom validity it would be difficult to identify that it was a symptom validity test and widely

available as a useful clinical tool. Furthermore, if the BVRT is validated as a symptom validity test and does not correlate with tests such as the TOMM, it could be used alongside the TOMM to increase diagnostic accuracy.

The use of more than one well validated symptom validity test has been shown to improve diagnostic accuracy (Victor et al., 2009) and is endorsed by the BPS (2009). However, the positive predictive accuracy of two tests which are highly correlated is no better than one and may be highly misleading (Rosenfeld, Sands & Van Gorp, 2000). Therefore the present study will examine the relationship between: Errors on the Benton Visual Retention Test; scores on the Reliable Digit Span- Revised (RDS-R; Young, Sawyer, Roper, & Baughman, 2012); and the recognition trial of the Logical Memory subtest of the Wechsler Memory Scale IV (Wechsler, 2010). If these tests are highly correlated, it would suggest that they should only be used independently.

A further aim of the present study is to examine the relationship between participants' severity of injury and their performance on symptom validity tests. The relationship between injury severity and symptom validity test results has been investigated previously with controversial findings: Green and Iverson (2001) examined the performance of 119 participants with head injuries ranging from mild to severe. They found an inverse (to expectation) relationship between head injury severity and performance on the Computerised Assessment of Response Bias (CARB; Conder, Allan & Cox, 1992). Severely injured participants performed better (i.e. tried harder) than those with milder injuries. The present study aims to examine whether or not the same pattern of test performance is observed in a UK clinical sample.

1.10. Hypotheses

1. That participants who are informed that they will be tested for effort will obtain higher scores than uninformed participants on the following measures:

    a) Test of Memory Malingering (TOMM; Tombaugh, 1995).
    b) Digit Span (DS) subtest of the Wechsler Adult Intelligence Scale (WAIS-IV-DS; Wechsler, 2010)
    c) Logical Memory (LM) subtest of the Wechsler Memory Scale (WMS-IV-LM; Wechsler, 2010)
    d) The Benton Visual Retention Test, number correct (BVRT; Sivan, 1992).
    e) Self reported anxiety levels

    Furthermore, it is hypothesised that the informed group will demonstrate lower BVRT 'error' scores than uninformed participants.

2. Litigating participants will score lower than non-litigating participants on the Test of Memory Malingering (TOMM; Tombaugh, 1995), Reliable Digit Span-Revised (RDS-R; Young, Sawyer, Roper, & Baughman, 2012) and the recognition trial of the Logical Memory subtest of the Wechsler Memory Scale IV (WMS-IV-LM-R; Wechsler, 2010). It is also hypothesised that litigating participants will make more distortion errors than non litigators who will make more omission errors on the Benton Visual Retention Test (BVRT; Sivan, 1992).

3. Length of Post Traumatic Amnesia will be positively correlated with TOMM trial II scores.

4. Participants with neuroimaging evidence of brain injury will obtain higher TOMM trial II scores than those without neuroimaging evidence of brain injury.

5. There will be a positive correlation between participants' scores on the following tests: Reliable Digit Span - Revised (RDS-R; Young, Sawyer, Roper, & Baughman, 2012) and the recognition trial of the Logical Memory subtest of the Wechsler Memory Scale IV (WMS-IV-LM; Wechsler, 2010). It is also hypothesised that there will be a negative correlation between the aforementioned tests and the number of distortion errors made on the Benton Visual Retention Tests (BVRT; Sivan, 1992) with participants scoring lower on other measures of symptom validity making more distortion errors.

2. Method

2.1. Design

The study utilised a between participants design where participants were either 'informed' or 'uninformed' that they would be tested for effort. Three independent variables were explored: 'Informed' status, level of self- reported anxiety and level of self- reported depression. Two dependant variables were explored: Symptom validity test performance and memory test performance.

2.2. Participants

21 Participants meeting the inclusion criteria were sampled from consecutive referrals to three UK neuropsychological assessment services between the months of January and July, 2012.

2.2.1. Inclusion and exclusion criteria

Participants were included if they had an acquired brain injury and were initially excluded if they had a degenerative or congenital condition or were suspected of having such. However, due to difficulty recruiting from the ABI population (n16), the inclusion criteria were broadened to include those with degenerative or congenital conditions, facilitating the recruitment of an additional five participants. They were aged over 18, spoke English as a first language and were judged by the assessing Neuropsychologist to have the capacity to consent to taking part in the research. Participants were excluded if they had a documented history of learning disability or previous diagnosis of dementia.

2.2.2. Context of assessment

Two of the neuropsychological assessment services were provided within the National Health Service (NHS), one based in the East Midlands and the other in the South East, UK. Participants were typically referred to the NHS services to assess the nature and extent of functional impairment or provide differential diagnoses. The resulting reports were used to guide decisions regarding treatment, care needs and diagnosis; although they may have also been used as evidence in litigation this would not have been the primary concern of the clinician. In contrast, the third service was a private neuropsychological assessment service specialising in medico-legal cases on behalf of the plaintiff. In this service the primary purpose of the assessment was to provide evidence regarding cognitive functioning after traumatic brain injury for use in litigation.

2.3. Measures

Participants' scores on symptom validity or effort tests were operationalised as scores on the Test of Memory Malingering (TOMM; Tombaugh, 1996). Participant performance on memory tests was operationalised as performance on a test of recall for verbally presented information, a test of recall for visually presented information and a test of working memory. For the purposes of the present research trait anxiety and depression were operationalised as participants' scores on the Hospital Anxiety and Depression Scale (HADS; Snaith & Zigmond, 1994).

2.3.1. Test of Memory Malingering

The TOMM was chosen as it is easy to administer and widely available in UK neuropsychology services. It is also one of the most frequently used effort tests in UK clinical practice (McCarter et al., 2009). It consists of two learning trials where the

participant is presented with 50 line drawings shown for three seconds each. The participant is then shown two drawings at a time, one of which is the same as a drawing they have just been shown in the learning trials. They have to indicate which drawing they have seen before for 50 pairs. A delayed retention trial of the same format is delivered 15 minutes later. A score of 45 or less out of 50 on the second retention trial indicates suboptimal effort and would meet Slick, Sherman and Iverson's (1999) B2 criterion for test evidence of 'probable' malingering. A score significantly below chance ($p<0.05$) would meet the Slick, Sherman and Iverson (1999) B1 criterion for 'definite' malingering. The TOMM has been validated by testing on non-clinical, clinical, clinical 'at risk of malingering' and simulated malingering samples. In a study comparing student simulators and controls the TOMM cut-off score of 45/50 yielded sensitivity and specificity rates of 100% (Tombaugh, 1997). According to Tombaugh (1996) internal consistency is high (coefficient alpha: trial 1= .94; trial 2= .95; retention trial = .94). Moore and Donders (2004) examined the test results of 132 brain injured participants and found that the diagnostic agreement rate between the TOMM and another symptom validity test (the California Verbal Learning Test) was 89% (k=0.40, $p<0.01$), suggesting they measure similar constructs.

2.3.2. Logical Memory

The Logical Memory subtest of the Wechsler Memory Scale IV (WMS-IV; Wechsler, 2010) is a test of recall for contextually meaningful verbally presented information. Examinees are read two different stories and asked to recall each straight away and after a short delay. The test provides an optional delayed forced-choice recognition trial where examinees are asked to respond 'true' or 'false' to a series of statements about the stories. Extensive standardisation studies have shown the Logical Memory Subtest of the WMS-IV to have good split-half reliability (r = 0.82 for

immediate recall and 0.85 for delayed recall). Test-retest reliability (within two months) was 0.72 and 0.67 for immediate and delayed recall; this was lower than expected due to practice effects (Wechsler, 2010). Standardisation sample scores on the Logical Memory Subtest of the WMS-IV were highly correlated (r = .76; Wechsler, 2010) with their scores on the same subtest of the highly validated WMS-III (Wechsler, 1997). The recognition trial of the Logical Memory subtest presents clients with a forced choice test where those participants scoring significantly below chance would meet Slick, Sherman and Iverson's (1999) B1 criterion for definite malingering.

2.3.3. Digit Span

The Digit Span Subtest of the Wechsler Adult Intelligence Scale IV (WAIS-IV; Wechsler, 2010) is a test of working memory, which is the examinee's ability to hold and manipulate information in their mind. Examinees are verbally presented with a short series of digits and asked to repeat them back immediately, either forwards (in the digits forwards trial), or backwards (in the digits backwards trial). Both trials consist of eight items, each consisting of two series of digits which increase in length incrementally if the examinee successfully recalls at least one series of digits from each trial. Split half reliability is good (r =0.93). The WAIS-IV Digit Span Subtest correlates well (r =.74; Wechsler, 2010) with its well validated predecessor from the WAIS-III. A further benefit of the Digit Span Subtest is that while it has high face validity as a test of memory, it relies mostly on attention and is easier than it seems. It is rare for patients with severe brain injuries (Axelrod, Fichtenberg, Millis, & Wertheimer, 2006) to score in the borderline or impaired ranges. The Reliable Digit Span- Revised (RDS-R; Young, Sawyer, Roper & Baughman, 2012) is calculated by summing the longest set of digits achieved in both trials for the digits forwards, backwards and sequencing sets. In a validation study (Young, Sawyer, Roper & Baughman, 2012) the RDS-R accurately

classified 64.9% of suspected malingerers who had failed the Word Memory Test (Green, 2003).

2.3.4. Benton Visual Retention Test

The Benton Visual Retention Test (BVRT; Sivan, 1992) is a popular test of visual recall, constructional ability, visuospatial perception and visuomotor response. Examinees are presented with line drawings on 10 cards. They are given 10 seconds to look at the card then the card is removed and they are asked to reproduce the design(s) on a blank sheet of paper. Examinees' responses are scored for both 'correctness' and 'errors' and can be compared to the normative data for people of the same age and estimated pre-morbid ability. Inter-rater reliability coefficients have been found to be good at .96 for number correct and .97 for errors (Swan, Morrison & Eslinger, 1990). Retesting healthy participants on two occasions at six month intervals produced no significant difference between test scores (Lezak, 1982; Cited in Lezak, Howieson & Loring, 2004).

The error scores of participants will be of interest to the current study as this alone has been used to differentiate between depressed and dementia patients (La Rue, D'Elia, Clark, Spar, & Jarvik, 1986). Errors are rare in healthy subjects (M=1.40) under the age of 80 who mostly make distortion errors with occasional rotational errors and omissions (Eslinger, Pepin & Benton, 1988). Therefore participants making several errors of a perseverative, rotational or misplacing nature or consistently omitting items from one side of the page are likely to be exhibiting signs of a genuine brain injury. Furthermore some studies have found that participants asked to simulate head injury make more distortion errors than people with genuine head injuries who mainly make omission errors (Benton & Spreen, 1961). For this reason the BVRT will also be used as an indicator of organicity.

2.3.5. Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression Scale (HADS; Snaith & Zigmond, 1994) provides separate scores for levels of both anxiety and depression but is relatively insensitive to level of brain injury. The HADS is a 14-item self-report questionnaire which produces two separate scores, one for depression and one for anxiety, both scored out of 21. Silverstone (1994) used a cut-off point of 8 out of 21 for clinical significance of symptoms and reported sensitivity rates for anxiety and depression to be 100% for medical and 80% for psychiatric in-patients. Specificity rates for the HADS (Snaith & Zigmond, 1994) were found to be 73% for medical patients (Silverstone, 1994). Silverstone (1994) notes that the higher the score the more sensitive and specific the HADS (Snaith & Zigmond, 1994) becomes. The HADS was selected because of its ease of administration and wide availability among services.

In order to measure subjective 'state' anxiety, before, during, and after testing, participants were asked to rate how anxious they felt on a Likert scale of 1 (not at all anxious) to 5 (very anxious).

2.4. Procedure

All patients referred to the collaborating services between the months of January and June 2012 were judged against the inclusion criteria and, as appropriate, invited to participate in the research. Participants were assigned to 'informed' or 'uniformed' conditions according to an *a priori* block randomisation procedure where both the condition assignment and block size was randomised. This procedure used random numbers generated by a programme on the website www.sealedenvelope.com, through which randomisation tables for each service were generated. The collaborating

clinicians assigned participants to conditions in referral order according to their randomisation table.

Potential participants were read an information sheet explaining that the study was examining factors affecting test results without suggesting that the focus of the study was effort testing (see appendix A). The information sheet stated that by agreeing to take part in the study the client was agreeing to have their test results and information about their condition shared with the chief investigator at the University of Leicester. Potential participants were informed that no resulting publications would contain identifiable personal information. Those in agreement with the terms in the information sheet were given a consent form to sign (see appendix B).

All participants were given standardised instructions; these were presented verbally, by the examiner and visually, printed on plain paper (in 'ariel' font size 14) for the participant to read. Participants in the 'uninformed' condition were presented with the following information:

I want to make sure that your test results are accurate.

Your test results will not be accurate if you get tired.

Your test results will not be accurate if you do not try your best.

Please try your best.

Please let me know if you get tired or need a break.

Participants in the 'informed' condition were presented with more information regarding effort testing:

I want to make sure that your test results are accurate.

Your test results will not be accurate if you get tired.

Your test results will not be accurate if you do not try your best.

An Effort Test measures how hard you are trying.

We routinely use Effort Tests.

I will give you an 'Effort Test' to measure how hard you are trying.

Please try your best. Please let me know if you get tired or need a break.

Participants were asked to repeat these instructions back to the examiner to ensure they had registered the information. No further information on effort tests was given and the effort test remained unidentified to the participants.

Prior to testing, the participants were asked to rate their state anxiety on a scale of one (not at all anxious) to five (panic). The collaborating clinicians proceeded with their assessments as they normally would, using tests that were appropriate for the client in the order they deemed appropriate. For the purposes of the present research tests included: the Logical Memory subtest of the WMS-IV (Wechsler, 2010), the Digit Span subtest of the WAIS IV (Wechsler, 2010), the Benton Visual Retention Test (BVRT; Sivan, 1992), the TOMM (Tombaugh, 1996) and the HADS (Snaith & Zigmond, 1994). In services assessing clients over more than one testing session these tests were administered at the initial testing session with the TOMM administered around the midway time point of the session. The clinicians asked participants to rate their anxiety again, directly before administering the TOMM and at the end of the testing session.

On completing each assessment the clinicians asked participants whether they were currently involved in any litigation regarding their brain injury or condition and

whether they have previously claimed compensation for this, or any other injury or condition.

Following the assessment the participants were debriefed and the full purpose of the study was explained to them verbally and on an information sheet (see appendix C.). The participants were reminded of their right to withdraw their data from the study. In order to preserve test security the TOMM (Tombaugh, 1996) remained unidentified to the participants.

The collaborating clinicians calculated their participants' test scores on the above measures, stated the nature of the participants head injury or condition, length of PTA (if appropriate) and commented on any neuroimaging evidence on a data collection form (see appendix D) which they returned to the chief investigator.

2.5. Ethical Issues

Ethical approval was sought and gained from Nottingham Research Ethics Committee-1, copies of correspondence with the committee can be found in appendix E.

2.6. Statistical Analyses

In the absence of previous research to indicate potential effect sizes, the required sample size was calculated on the assumption of a medium effect size (0.15). In order to detect a medium effect size (α=0.05, power 0.8), a total number of 92 participants were required.

3. Results

3.1. Introduction

In order to provide a coherent description of the data, analyses and findings, this section will start by contextualising the sample in terms of participant characteristics. The equivalence of experimental groups will be examined prior to testing of the primary hypotheses which will be explored thoroughly before engaging with and testing of secondary hypotheses. Finally, the findings of the study will be summarised.

3.2. Sample Characteristics

21 patients agreed to participate in the study, of whom 57% (n= 12) were male and 43% (n= 9) were female. The age of participants ranged from 21 to 63 years (m = 40.62; SD = 11.72) and years in formal education varied from11 to 18 (m = 12.71; SD = 2.21). Participants' presenting problems were categorised into the following groups:

- Traumatic Brain Injury (TBI): Included participants whose brain injuries were sustained through road traffic accidents, assaults or blows to the head.
- Acquired Brain Injury (ABI): Included participants who had survived one of the following: Stroke, Anoxia, Aneurysm or Encephalitis.
- Congenital Condition (CC): Included participants seeking neurological assessment due to Epilepsy or an Arteriovenous Malformation which had resulted in brain haemorrhage.
- Symptoms Only (SO): Included one client presenting with concerns regarding their cognitive functioning without any known brain injury or condition.

The demographics of the sample are displayed in table 3.

Table 3: Sample Characteristics

Diagnosis	N	Male	Age	PM IQ	Ed	Months Since Injury	PTA	Litigating
TBI	11	55%	39 (12)	96 (9)	13 (2)	23 (15)	22 (39)	63%
ABI	5	60%	50 (3)	97 (10)	12 (2)	71 (116)	N/A	0%
CC	4	50%	30 (9)	94 (17)	13 (2)	N/A	N/A	0%
SO	1	100 %	51	83	11	N/A	N/A	0%
Total	21	57%	41 (12)	95 (11)	13 (2)	38 (65)		33%

Number of Participants; Percent Male; Mean Age (SD); Mean Estimated Pre-morbid Full Scale Intelligence Quotient (SD), Mean Years in Formal Education (SD), Mean Months Since Injury (SD), Mean Length of Post Traumatic Amnesia in Days (SD) and percent seeking litigation at time of assessment arranged by diagnostic category.

3.3. Data Processing

The collaborating clinicians submitted their data collection forms (appendix D) to the chief investigator. Some of the data collection forms were returned incomplete: where possible, the missing data were sought and retrieved from the collaborating clinician. The final data set was complete with the exception of Hospital Anxiety and Depression Scale (HADS; Snaith & Zigmond, 1994) scores for one participant and Benton Visual Retention Test (BVRT; Sivan, 1992) scores for four participants. Raw data and standardised scores were entered into an SPSS database from which all analyses were run.

3.4. Group Equivalence

Participants were assigned to either ‘informed’ or ‘uninformed’ conditions by a process of block randomisation. The characteristics of the informed and uninformed groups were assessed for equivalence by visual comparison (see table 4).

Table 4. Comparison of Group Characteristics

Condition	N	Male	Age	PMIQ	Years in Education	Months Since Injury	PTA	Litigation
Informed	9	67%	47 (10)	94 (10)	11 (1)	14 (15)	9 (24)	33%
Uninformed	12	50%	36 (11)	96 (12)	14 (2)	57 (83)	14 (34)	33%

(Number of Participants; Percent male; Mean age (SD); Mean Pre-morbid Full Scale Intelligence Estimate (SD), Mean years in formal education (SD), Mean months since injury (SD), Mean length of Post Traumatic Amnesia in days (SD) and percent seeking litigation at time of assessment for 'informed' and 'uninformed' groups).

Further assessment of group equivalence was conducted by using the error and correct scores on the Benton Visual Retention Test (BVRT; Sivan, 1992) as an indicator of head injury and the Hospital Anxiety and Depression Scale (HADS; Snaith & Zigmond, 1994) as a measure of anxiety and depression. The descriptive statistics for these comparisons are displayed in table 5.

Table 5. Comparisons of Groups' Condition at Assessment

Condition	Reason for Assessment	BVRT Correct	BVRT Error	HADS Anxiety	HADS Depression
Informed	44% TBI 33% ABI 11% CC 11% SO	4 (2)	7 (2)	9 (5)	6 (4)
Uninformed	58% TBI 17% ABI 25% CC	5 (4)	7 (2)	10 (4)	9 (3)

Benton Visual Retention Test (BVRT;) Mean Correct (SD); Mean Error scores (SD); Mean Hospital Anxiety and Depression Scale Scores (HADS) for anxiety (SD) and depression (SD).

3.5. Characteristics of Dependent Variables

Prior to the use of inferential statistics, each dependent variable was examined for normality. Scores on both trials of the TOMM and the Logical Memory-Recognition test were not normally distributed, therefore violating the parametric assumptions.

Logical Memory I and II, Digit Span age scaled score, Reliable Digit Span, HADS scores and Benton Visual Retention error and correct scores were normally distributed (see appendix F).

In order to avoid entering correlated dependent variables into the planned Multivariate Analysis of Variance a series of correlations were performed using a non-parametric test to accommodate the TOMM's lack of normality. These indicated significant relationships between several dependant variables:

- TOMM trial II scores were correlated with: TOMM Trial I scores (tau-b = 0.602, $p<0.05$); Logical Memory I age scaled scores (tau-b = 0.473, $p<0.05$) and Logical Memory II age scaled scores (tau-b = 0.527, $p<0.001$).
- Digit Span age scaled scores were correlated with: Logical Memory I age scaled scores (tau-b = 0.415, $p< 0.05$) and Logical Memory II age scaled scores (tau-b = 0.388, $p<0.05$)
- The Benton Visual Retention Test 'error' score was correlated with its 'correct' score (tau-b = -0.841, $p<0.001$)

3.6. Testing Primary Hypotheses: The effect of informed consent

It was hypothesised that 'informed' participants would obtain higher scores than 'uninformed' participants on the following measures:

f) Test of Memory Malingering (TOMM; Tombaugh, 1995).

g) Digit Span (DS) subtest of the Wechsler Adult Intelligence Scale (WAIS-IV-DS; Wechsler, 2010)

h) Logical Memory (LM) subtest of the Wechsler Memory Scale (WMS-IV-LM; Wechsler, 2010)

i) The Benton Visual Retention Test, number correct (BVRT; Sivan, 1992).

j) Self reported anxiety levels

It was further hypothesised that the informed group would demonstrate lower BVRT 'error' scores than uninformed participants.

Table 6. Summary of Symptom Validity and Memory Test Results for Informed and Uninformed Groups.

Test	Informed	Uninformed
TOMM I	46.44 (6.06)	47.25 (5.36)
TOMM II	49.67 (1)	48.58 (4.32)
Logical Memory I age scaled score	8 (3.94)	7.5 (4.21)
Logical Memory II age scaled score	7.78 (3.93)	7.42 (4.32)
Logical Memory retention total score	25.33 (3.84)	22.83 (4.93)
Digit Span age scaled score	8.89 (4.48)	9.25 (3.17)
Benton Visual Retention Test Number Correct	7.25 (1.67)	7 (2.24)
Benton Visual Retention Test Error Score	3.62 (2.39)	5 (4.56)
Anxiety before testing commenced	2.44 (1.01)	2 (1.34)
Anxiety directly before TOMM	2.56 (1.33)	2.67 (1.56)
Anxiety at the end of testing	1.88 (1.05)	2.33 (1.30)

Mean (SD) TOMM, Memory Test and Self Reported Anxiety Scores for Informed and Uninformed Participants.

Table six shows similar mean scores on each dependent variable between informed and control groups. The observed standard deviations within each group suggest considerable overlap in the distributions of scores between groups and, therefore, that there is unlikely to be a significant difference between groups.

In order to test whether there was any significant difference in TOMM scores between groups a Mann-Whitney U Test was conducted. This revealed no significant difference between informed and uninformed groups (see appendix F).

In order to test whether there were significant differences between groups on the Logical Memory II test, BVRT correct score and Digit Span Test these were entered as Dependent variables into a Multivariate Analysis of Variance (MANOVA). To avoid entering correlated dependant variables the Logical Memory I and error score on the BVRT were omitted from the analysis.

There was no significant effect of information (informed or uninformed) on the combined dependent variable ($F_{(3,13)} = 0.373$, ns). Analysis of each individual dependent variable found no significant effect of information on BVRT correct scores ($F_{(1,13)} = 0.067$, ns) Logical Memory II ($F_{(1,13)} = 0.027$, ns) or Digit Span age scaled scores ($F_{(1,13)} = 0.83$, ns).

In order to assess whether non equivalence of groups contributed to the null findings pre-test differences in trait anxiety and depression, age, years in education, pre-morbid IQ, months since injury and length of PTA were compared using a MANOVA. Anxiety and depression scores on the Hospital Anxiety and Depression Scale (HADS; Snaith & Zigmond, 1994) were correlated (tau-b=0.407, $p<0.05$), therefore anxiety scores were omitted from and depression scores were included in the analysis. Pre-morbid IQ estimates (Test of Pre-Morbid Functioning; Wechsler, 2010) and years in education were also correlated (tau-b= 0.579, $p<0.001$), as years in education violated the parametric assumptions (Levene's $p> 0.05$) it was replaced with estimated pre-morbid IQ in the analysis.

There was no significant difference between control and informed groups on the combined measures of equivalence ($F_{(5,8)} = 2.659$, ns, Wilks' Lambda = 0.376; partial eta squared = .624). Full details of this analysis are available in appendix G.

3.6.1. State anxiety

Due to the non-parametric nature of Likert scale ratings differences between informed and uninformed participant levels of state anxiety were excluded from MANOVA analysis. A descriptive summary of participants self reported anxiety before, during and after testing is displayed in figure 1.

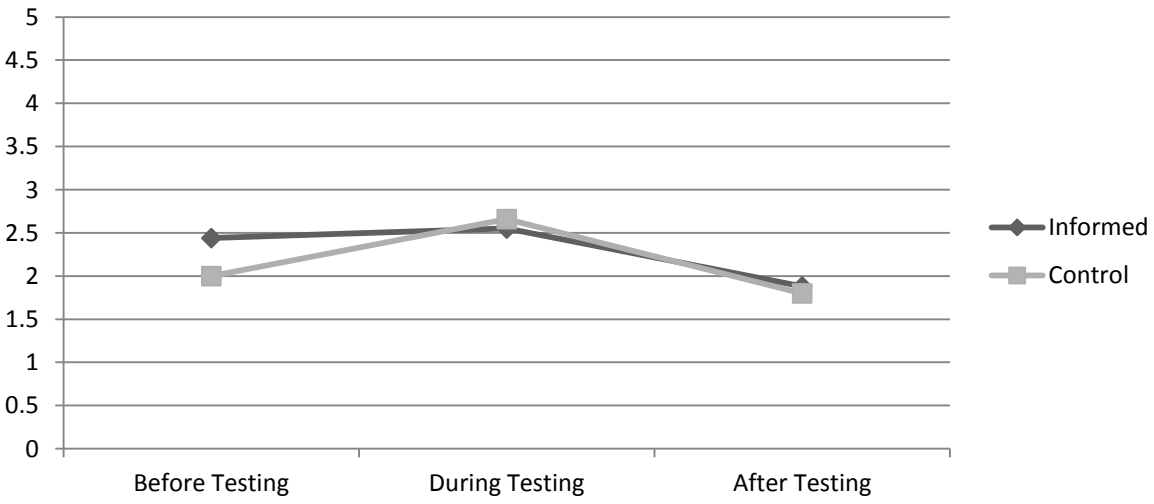


Figure 1. Self Reported Anxiety Levels in Informed and Control Participants

Figure 1. Shows participant average self report anxiety levels for informed and uninformed groups on a scale of: 0 (no anxiety) to 5 (panic). Visual inspection of the data revealed a trend for all participants to report a slight increase in anxiety during testing (immediately before the TOMM) and a decrease in anxiety at the end of testing. Mann-Whitney U tests revealed no significant differences in self-reported anxiety between informed and uninformed participants before, during or after testing.

3.7. Litigation status and SVT results

It was hypothesised that there would be a difference between litigating and non-litigating participant scores on the Test of Memory Malingering (TOMM; Tombaugh, 1996), Reliable Digit Span- Revised (RDS-R; Young, Sawyer, Roper, & Baughman, 2012) and the recognition trial of the Logical Memory subtest of the Wechsler Memory Scale IV (WMS-IV-LM-R; Wechsler, 2010). It was also hypothesised that litigating participants would make more distortion errors than non litigators who would make more omission errors on the Benton Visual Retention Test (BVRT; Sivan, 1992).

Table 7. Mean and Standard Deviation SVT Scores for Litigating and Non Litigating Participants

	Litigating	Non Litigating
Logical Memory Recognition Total Score	24.86 (5.24)	23.43 (4.33)
Reliable Digit Span	17.14 (5.11)	15.93 (3.32)
TOMM Trial II	47.86 (5.66)	49.64 (0.93)
Benton Visual Retention Test- Error Score	3.4 (2.19)	4.75 (4.13)
BVRT – Omissions	0.2 (0.44)	0.92 (1.78)
BVRT- Rotations	0.6 (0.89)	1.08 (0.1)
BVRT- Distortions	1.8 (2.04)	1.42 (1.93)
BVRT- Perseverations	0.4 (0.89)	0.42 (0.79)
BVRT- Misplacements	0.2 (0.45)	0.83 (1.7)
BVRT- Size	0.2 (0.45)	0 (0)

Visual examination of the data presented in table seven suggests that there was no significant difference between the performance of litigating and non litigating participants. Due to the small size of the litigating sample (7, with missing BVRT data

for 2) use of ANOVA was deemed inappropriate and Mann-Whitney U tests were conducted, these confirmed that there was no significant difference on any measure of symptom validity between litigating and non-litigating participants. On examination of the raw data (see appendix H), one litigating participant scored 35/50 on the TOMM, bringing the group average down. All other litigating participants scored 50/50. In order to gain a more qualitative understanding of participant SVT performance scores below standard cut-offs are presented in table eight.

Table 8. Frequency of SVT Failure in Litigating and Non Litigating Groups

	N	Failing TOMM II	Failing RDS-R	Below Chance performance on LMR	Failing more than one measure of Symptom Validity
Litigating	7	1	1	0	1
Non Litigating	14	0	1	0	0

Only one participant scored below standard cut offs on the TOMM II (below 45/50), the same participant scored below cut offs for the RDS-R (below 11) but did not score significantly below chance on the recognition trial of the LMR: they did, however, score at chance level (15/30).

3.8. The relationship between injury severity and scores on the TOMM

It was hypothesised that for participants who had survived a traumatic brain injury (n= 11) there would be a negative correlation between length of Post Traumatic Amnesia (in days) and their scores on trial II of the Test of Memory Malingering (Tombaugh; 1996). No significant relationship was detected (tau-b= .032, p = ns). It was also hypothesised that within the whole sample, participants with neuroimaging

evidence of brain injury (n= 14) would achieve higher scores on the TOMM than participants without neuroimaging evidence of brain injury (n= 7). A Mann-Whitney U test did not detect any significant difference in TOMM trial II scores between participants with and without neuroimaging evidence of brain injury (see appendix I).

3.9. The relationship between embedded measures of Symptom Validity

It was hypothesised that there would be a positive relationship between participants' scores on the following tests: Reliable Digit Span - Revised (RDS-R; Young, Sawyer, Roper, & Baughman, 2012) and the recognition trial of the Logical Memory subtest of the Wechsler Memory Scale IV (WMS-IV-LM; Wechsler, 2010). It was also hypothesised that there would be a negative correlation between the aforementioned tests and the number of distortion errors made on the Benton Visual Retention Tests (BVRT; Sivan, 1992) with participants scoring lower on other measures of symptom validity making more distortion errors.

Participant scores on measures of symptom validity were entered into SPSS and analysed using Kendall's tau-b. In order to adjust for multiple comparisons correlations were only considered significant if $p<0.01$. The results are presented in table nine.

Table 9. The Relationship between Scores on Different Symptom Validity Measures

	LM-R	**RDS-R**	**BVRT Distortions**
LM-R			
RDSR	.273ns		
BVRT Distortions	.044ns	-.390ns	

There was no evidence of significant relationships in scores between LM-R, RDS-R and BVRT distortion scores. Full details of this analysis are available in appendix J.

3.10. Summary of Findings

In the context of a limited sample size, there were no significant differences between informed and uninformed groups on the Test of Memory Malingering, trial II (TOMM; Tombaugh, 1996), the Logical Memory subtest of the Wechsler Memory Scale (WMS-IV; Wechsler, 2010), and the Benton Visual Retention Test (BVRT; Sivan, 1992). Despite predictions there was no significant difference between informed and uninformed groups in their self-ratings of anxiety, nor its fluctuation throughout the testing session.

No significant differences were observed between litigating and non litigating participants on the Recognition Trial of the Logical Memory Test (WMS-IV; Wechsler, 2010), Reliable Digit Span-Revised (RDS-R; Young, Sawyer, Roper, & Baughman, 2012), Benton Visual Retention Test (Sivan, 1992) or on the TOMM trial II.

One litigating participant failed trial II of the TOMM, scoring 35/50; the same participant also scored below the cut-offs for RDS-R but not below chance on the LM-R, scoring exactly 50%. One non-litigating participant failed the RDS-R, scoring below cut offs but passed both the other SVTs.

Unexpectedly, correlations between symptom validity measures were low, absent or undetected.

4. Discussion

4.1. Summary

The present research aimed to determine whether informing neuropsychology examinees that they will be assessed for effort affects their symptom validity and memory test results. While the British Psychological Society (BPS, 2009) advises that clients should be informed that they will be tested for effort, symptom validity tests are designed and normed on uninformed samples. Therefore, if informing clients that they will be tested for effort affects any of their test results it would invalidate the assessment (Boone, 2007).

4.1.1. Primary research hypotheses

It was hypothesised that participants who were informed they would be tested for effort would achieve higher scores on symptom validity and memory tests than uninformed participants. In order to test this hypothesis participants were recruited from a neuropsychology outpatient population as they presented for neuropsychological assessment. Patients who agreed to participate were randomly assigned to either an informed or an uninformed condition and underwent neuropsychological testing, including the Test of Memory Malingering (TOMM; Tombaugh, 1996) and other measures of symptom validity embedded within memory tests.

Due to recruitment difficulties the final sample size was not sufficient for adequate statistical power. There was no detected significant difference between informed and uninformed groups on symptom validity, self reported anxiety levels or memory test results.

4.1.2. Secondary research hypotheses

The BPS (2009) acknowledges concerns that informing patients they will be tested for effort may raise their levels of state anxiety. The present study sought to investigate whether self-report anxiety levels were higher in the informed group when compared to the uninformed group. There was no observed difference in self report anxiety levels between the groups at either pretesting, pre-effort testing or end of testing time points.

Previous research has indicated that the Benton Visual Retention Test (BVRT; Sivan, 1992) may have utility as a non-forced choice symptom validity test (Benton & Spreen, 1961). The present research employed a differential prevalence design to assess whether frequency of distortion errors on the BVRT could be used to differentiate between litigating and non litigating participants. There was no observed difference between litigating and non litigating groups on The BVRT, or indeed, other measures of symptom validity.

There is an increasing interest in the development of symptom validity tests which are not highly correlated with other symptom validity tests. This is because while the use of more than one well validated measure of symptom validity can increase diagnostic accuracy (Victor et al., 2009) and is endorsed by the BPS (2009). The positive predictive accuracy of two tests which are highly correlated is no better than one. Furthermore, if a clinician assumes that positive predictive accuracy is cumulative, results may be highly misleading (Rosenfeld, Sands & Van Gorp, 2000). Therefore the present study examined the relationship between: Errors on the Benton Visual Retention Test (Sivan, 1992); scores on the Reliable Digit Span- Revised (RDS-R; Young, Sawyer, Roper, & Baughman, 2012); and the recognition trial of the Logical Memory subtest of the Wechsler Memory Scale IV (Wechsler, 2010). There was no evidence of

a significant relationship between scores on these embedded measures of symptom validity, however, due to the small sample size this could be due to low statistical power rather than the absence of a relationship.

In response to previous studies (e.g. Green & Iverson, 2001) demonstrating a positive relationship between severity of injury and effort test results, the present study sought to examine the relationship between severity of injury and effort test results. There were no observed differences in TOMM scores between participants with and without neuroimaging evidence of brain injury. For participants who had experienced a period of Post Traumatic Amnesia (PTA) there was no detected relationship between length of PTA and TOMM scores.

4.2. Theoretical and Practical Value of Research

4.2.1. Interpretation of findings

The present study failed to achieve adequate statistical power to address primary and secondary research questions resulting in a high risk of type II error in any interpretation of the results.

4.2.2. Comparability with previous research

Previous research has been stimulated by suspicions that North American Attorneys may warn or even coach litigants regarding effort testing. In contrast, the present study was stimulated by concerns that following BPS (2009) guidance and informing clients that they will be tested for effort contradicts test manual instructions. As such, the present study aimed to examine differences in test performance between participants who were either informed or uninformed they would be tested for effort. The act of informing a participant represented a more subtle departure from test manual guidance than the frank warnings presented in previous research (e.g. King & Sulivan,

2009). While previous research has endorsed the use of students simulating head injury the present study sought to increase ecological validity and relevance to UK clinical practice through the use of a clinical sample. In summary, the present research was more relevant to UK clinical practice than previous research and attempted to be more ecologically valid through the use of a 'real life' clinical sample. Unfortunately, the use of a clinical sample raised recruitment issues which have severely impacted on the statistical power of the present study.

4.2.3. Theoretical strengths

A particular strength of the present research is the theory to practice link made in asking the main research question: For psychometric test results to be valid, it is a necessary (but not sufficient) condition that the clinical sample is exposed to similar conditions as the normative sample. It is difficult to predict to what extent differences in instructions and methods of delivery will affect test performance and validity. However, previous research has suggested that extra information or warnings about tests can impact on test performance. If neuropsychological practice is to be both evidence based and ethical any differences in the test results of informed and uninformed clinical patients needs to be identified and investigated.

Whilst there was no observed relationship between the Benton Visual Retention Test (Sivan, 1992) and other measures of symptom validity, repetition of the analyses with a larger sample may raise theoretically challenging issues. If the BVRT could reliably distinguish between litigating and non litigating groups and was highly correlated with the Test of Memory Malingering (TOMM; Tombaugh, 1996) it could be validated as a test of symptom validity. However, as a symptom validity test (SVT) highly correlated with other symptom validity tests it would not be appropriate to use it alongside other SVTs, limiting its clinical utility. Alternatively, should the BVRT

reliably distinguish between litigating and non litigating groups, but not correlate with the TOMM, it would suggest it is measuring a different construct. Such an outcome seems unlikely, given that the TOMM is also likely to distinguish between litigating and non litigating groups probably resulting in a correlation between the two. However, this thought experiment raises further issues regarding what two unrelated but validated measures of symptom validity might actually be measuring.

4.2.4. Methodological strengths

The use of multiple sites, clinicians and standardised instructions for both informed and uninformed conditions minimised the risk of clinician or site-related extraneous variables unduly influencing the data. Block randomisation procedures allowed the informed and uninformed group sizes to remain approximately equal throughout data collection without introducing any bias into condition allocation. While the use of multiple data collection sites and clinicians was designed to aid the achievement of the required sample size, it also enabled greater heterogeneity within the obtained sample, making it more likely to be representative of neuropsychology clients within the UK.

4.2.5. Methodological limitations

During the six months of data collection it became clear that the neuropsychology assessment services were finding it difficult to recruit patients to participate in the research. One of the reasons posited for these problems was the narrow (i.e. Traumatic Brain Injury only) inclusion criteria. In response to concerns from the collaborating clinicians the inclusion criteria were widened to permit inviting patients with acquired, congenital or degenerative conditions to participate. Whilst the widening of the inclusion criteria may have made the sample more representative of patients presenting for neuropsychological assessment in the UK it introduced a high

level of heterogeneity in cognitive ability between participants. This high level of variability in test results between participants may have decreased the study's statistical power to detect a small or medium sized effect for information between groups. Furthermore, the addition of participants with acquired or congenital conditions and no external incentive to perform poorly lowered the risk of poor TOMM performance within the sample. The majority of the sample scored 50/50 on trial II of the TOMM, resulting in very little variability in scores and making any analysis of differences between or within groups difficult.

A further recruitment issue raised by participating services was the increased workload produced by the inclusion of additional tests in their usual battery. Collaborating clinicians discussed finding it difficult to find the time to engage in the research process due to work load and other commitments. In response to these pressures two of the four services were offered the support of a Trainee Clinical Psychologist to test participants one day per week.

Despite effort to make reasonable adjustments, the final sample size was much smaller than originally anticipated, resulting in a lack of statistical power. Recruitment difficulties were, in part, due to inappropriate referrals; these included patients who did not have the mental capacity to consent to participation and those who were judged to be unable to attend to testing for an hour. One service in particular favoured only using tests appropriate to a referral question, rarely utilising whole test batteries. This made it difficult for the clinicians to justify subjecting their patients to additional testing.

Attitudes towards the use of Symptom Validity Tests (SVTs) varied within and between services: Two of the four participating services used SVTs as part of a routine battery of tests while the remaining services only used SVTs if they suspected

malingering. Despite BPS guidelines encouraging the use of SVTs with all client groups, one neuropsychologist was ethically opposed to the use of SVTs with people who had clear evidence of stroke and no external incentive to malinger. Indeed, in such a context, if the client were to 'fail' the TOMM, they would not meet the Slick, Sherman and Iverson (1999) criteria for malingering on the grounds of the absence of external incentive. However, TOMM results would be endorsed by the BPS on the grounds that any evidence of response bias would inform the interpretation of other neuropsychological test results.

Another recruitment issue arose when one clinician asked to include a client whom he suspected might be malingering; he felt it was clinically important that this client was assigned to the 'informed condition'. As a result of this clinical preference the client was subsequently excluded from the study to prevent introducing bias in condition assignment.

4.3. Research Implications

The present study sought to investigate whether informing clients that they would be tested for effort affected their symptom validity test results. This remains an important and unanswered question. Data collection is ongoing with the aim of achieving adequate statistical power to examine for differences between informed and uninformed groups. Difficulties with data collection raise further, previously unforeseen research questions that may have proved useful precursors to the present study:

1. Under what circumstances do UK Neuropsychologists employ Symptom Validity Tests?

2. Do Neuropsychologists inform their clients that they will be tested for effort?
3. If Neuropsychologists inform their clients they will be tested for effort what do they say?
4. How do Neuropsychologists interpret symptom validity test failure?
5. If Neuropsychologists inform their clients they will be tested for effort, does this alter their interpretation of test results

While the present study did not attempt to provide estimates of the base rates of malingering in UK litigating and non-litigating populations this was observed to be an area for future research to investigate. The positive predictive accuracy of symptom validity tests such as the Test of Memory Malingering (TOMM; Tombaugh, 1996) in diagnosing definite or probable malingering (Slick, Sherman & Iverson, 1999) are dependent on the base rate of malingering within the population sampled. Unfortunately there is no published data on the likely base rates of malingered neurocognitive dysfunction in the UK (BPS, 2009). A recent review (Nicholls, 2012) of four North American studies yielded 503 litigating, traumatic brain injury participants of whom 24.55% were identified as either probably or definitely malingering according to the Slick, Sherman and Iverson criteria (1999). Meta analysis of The TOMM's sensitivity and specificity suggest they are 65 % and 94% respectively (Sollman & Berry, 2011). With a (North American) base rate of 25% positive predictive accuracy for the TOMM would be 82%, however, the base rate of 25% is derived from a 'high risk' population. If the base rate were lower as could be expected in a UK non-litigating sample, positive predictive accuracy would fall accordingly. If UK base rates were as low as 10% the TOMM's positive predictive accuracy would drop to 55%, meaning that there would be a 45% likelihood that a positive result on the TOMM was inaccurate. Therefore, without

reliable UK base rate estimates symptom validity tests such as the TOMM may have limited clinical utility.

4.4. Clinical Implications

An unintended consequence of the present study was the observation of a variety of different attitudes towards both the use of symptom validity tests and the act of informing clients they will be tested for effort. There appeared to be an inconsistency in the use of SVTs both within and between neuropsychology assessment services with some services using the tests as standard practice and others reserving their use for cases they suspected of malingering. Opinions and practices regarding informing clients they would be tested for effort also varied, with some clinicians preferring to follow BPS guidance and others, concerned they might affect the validity of the test, opting to adhere to the test manual and disregarding BPS advice. The act of raising awareness of the present research question has in and of itself proven clinically useful in stimulating discussion and reflection within clinical teams who had not previously considered these issues. Given that, without adequate base rate data, the validity of symptom validity tests is potentially confounded and may be further confounded by following BPS advice clinicians should exercise extreme caution in any interpretation of SVT results.

4.5. Conclusion

The present study was conducted to investigate whether informing patients they would be assessed for effort affects their symptom validity test results. No differences were observed between informed and uninformed groups in scores on the Test of Memory Malingering (TOMM; Tombaugh, 1996), embedded symptom validity measures or memory tests.

Secondary research questions were also addressed: There was no difference between litigating and non litigating groups in scores on the Benton Visual Retention Test (BVRT; Sivan, 1992) or any other test of symptom validity. Unexpectedly, there were no significant correlations between embedded symptom validity measures. These findings must be understood in the context of low statistical power where any interpretations of these data risk making a type II error. While no difference was observed between groups, the present study's statistical power was so limited it was not possible to determine whether a lack of difference between groups was due to the absence of an effect or the failure detect an effect.

5. References

Axelrod, B. N., Fichtenberg, N. L., Millis, S. R., & Wertheimer, J. C. (2006). Detecting incomplete effort with digit span from the Wechsler Adult Intelligence Scale- third edition. *The Clinical Neuropsychologist, 20,* (3). 513-523.

Benton, A.L., & Spreen, O. (1961). Visual memory test: The simulation of mental incompetence. *Archives of General Psychiatry, 4,* 79-83.

Boone, K. B. (2007). *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective.* New York: Guilford Press.

British Psychological Society; Professional Practice Board (2009). *Assessment of Effort in Clinical Testing of Cognitive Functioning for Adults.* Leicester, UK: British Psychological Society.

Coleman R.D., Rapport L.J., Millis S.R., Ricker J.H., & Farchione T.J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology, 20,* (2), 201-210.

Conder, R., Allen, L., & Cox, D. (1992). *Manual for the Computerized Assessment of Response Bias.* Durham, NC: CogShell.

Dunn, T. M., Shear, P. K., Howe, S., & Ris, M. D. (2003). Detecting neuropsychological malingering: Effects of coaching and information. *Archives of Clinical Neuropsychology, 18,* 121-134.

Eslinger, P. J., Pepin, L., & Benton, A. L. (1988). Different patterns of visual memory errors occur with aging and dementia. *Journal of Clinical and Experimental Neuropsychology, 10,* 60-61.

Green, P. (2009). Spoiled for choice: Making comparisons between forced-choice effort tests. In K. B. Boone (Ed.), *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective* (pp 50-77). New York: Guildford Press.

Green, P., & Iverson, G. (2001). Validation of the Computerized Assessment of Response Bias in litigating patients with head injuries. *The Clinical Neuropsychologist, 15* (4), 492-497.

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.

Gunstad, J., & Suhr, J. A. (2001). Efficacy of the full and abbreviated forms of the Portland Digit Recognition Test: Vulnerability to coaching. *The Clinical Neuropsychologist, 15,* 397-404.

Jelicic, M., Ceunen, E., Peters, M. J. V., & Mercklebach, J. (2011). Detecting coached feigning using the Test of Memory Malingering (TOMM) and the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology, 67,* (9) 850-855.

Jelicic, M., Merckelbach, H., Candel, I., & Geraerts, E. (2007). Detection of feigned cognitive dysfunction using special malinger tests: A simulation study in naive and coached malingerers. *International Journal of Neuroscience, 117,* 1185-1192.

Johnson, J. L., & Lesniak-Karpiak, K. (1997). The effect of malingering on memory and motor tasks in college samples. *Archives of Clinical Neuropsychology, 12,* (3) 231-238.

King, J., & Sullivan, K. A. (2009). Deterring malingered psychopathology: The effect of warning simulating malingerers. Behavioural Sciences and the Law, 27, 35-49.

La Rue, A., D'Elia, L. F., Clark, E. O., Spar, J. E., & Jarvik, L. F. (1986). Clinical tests of memory in dementia, depression, and healthy aging. *Psychology and Aging, 1,* (1), 69-77.

Larabee, G. J. (2007). *Assessment of Malingered Neuropsychological Deficits.* New York: Oxford University Press.

Lezak, M. D., Howieson, D. B., Loring, D. W. (2004). *Neuropsychological Assessment: Fourth Edition.* New York: Oxford University Press.

Martin, R., Bolter, J., Todd, M., Gouvier, W., & Niccolls, R. (1993). Effects of sophistication and motivation on detection of malingered memory performance using computerized forced-choice tasks. *Journal of Clinical and Experimental Neuropsychology, 15,* 867-880.

McCarter, R. J., Walton, N. H., Brooks, D. N., & Powell, G. E. (2009). Effort testing in contemporary UK neuropsychological practice. *The Clinical Neuropsychologist, 23,* 1050-1066.

Moore, B. A., & Donders, J. (2004). Predictors of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury, 18,* (10), 975-984.

Moss, A., Jones, C., Folkias, D. & Quinn, D. (2003). The mediating effects of effort upon the relationship between head injury severity and cognitive functioning. *Brain Injury, 17,* (5) 377-387.

Nitch, S. R., & Glassmire, D. M. (2007). Non-forced-choice measures to detect non-credible cognitive performance. In K. B. Boone (Ed.), *Assessment of Feigned Cognitive*

Impairment: A Neuropsychological Perspective (pp 78-102). New York: Guildford Press.

Pankratz, L. (1988). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception.* New York: Guildford Press.

Powell, M. R., Gfeller, J. D., Oliveri, M. V., Stanton, S. & Hendricks, B. (2004). The Rey AVLT serial position effect: A useful indicator of symptom exaggeration? *The Clinical Neuropsychologist, 18,* (3) 465-476.

Rose, F.E., Hall, S., Szalda-Petree, A. D., & Bach, P. J. (1998). A comparison of four tests of malingering and the effects of coaching. *Archives of Clinical Neuropsychology, 13,* 349-363.

Rosenfeld, B., Sands, S. A., & Van Gorp, W.G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology, 15,* (4) 349-359.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22,* 213-223.

Silverstone, P. H. (1994). Poor efficacy of the Hospital Anxiety and Depression Scale in the diagnosis of major depressive disorder in both medical and psychiatric patients. *Journal of Psychosomatic Research, 38,* 441-450.

Sivan, A. B. (1992). *Benton Visual Retention Test* (5th Ed.). San Antonio, TX: The Psychological Corporation.

Snaith, R. P., & Zigmond, A. S. (1994). *Hospital Anxiety and Depression Scale.* London, UK; nferNelson.

Suhr, J. A. (2002). Malingering, coaching and the serial position effect. *Archives of Clinical Neuropsychology, 12,* 69-77.

Suhr, J. A., & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology, 15*, 415-424.

Schenk, K., & Sullivan, K. A. (2010). Do warnings deter rather than produce more sophisticated malingering? *Journal of Clinical and Experimental Neuropsychology, 32,* (7) 752-762

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13,* 545-561.

Sullivan, K., Deffenti, C., & Keane, B. (2002). Malingering on the RAVLT part II. Detection strategies. *Archives of Clinical Neuropscyhology, 12,* 223-233.

Sullivan, K., Keane, B., & Deffenti, C. (2001). Malingering on the RAVLT part I. Deterrence strategies. *Archives of Clinical Neuropsychology, 16,* 627-641.

Sullivan, K., & Richer, C. (2002). Malingering on subjective complaint tasks an exploration of the deterrent effects of warning. *Archives of Clinical Neuropsychology, 17,* 691-708.

Sullivan, K., Lange, R., & Dawes, S. (2005). Malingering base rates and detection methods in Australia. *Journal of Forensic Neuropsychology* 4, 49-70.

Swan, G. E., Morrison, E., & Eslinger, P. J. (1990). Interrator agreement on the Benton Visual Retention Test. *The Clinical Neuropsychologist, 4,* 37-44.

Tombaugh, T. N. (1996). *Test of Memory; TOMM.* North Tonawanda, NY: Multi-Health Systems Inc.

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

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J. & Ziegler, E. A. (2009). Interpreting the Meaning of Multiple Symptom Validity Test Failure. *The Clinical Neuropsychologist,23*, 297-313.

Wechsler, D. (2010). *Wechsler Adult Intelligence Scale- Fourth UK Edition.* San Antonio, TX: Pearson Assessment.

Wechsler, D. (2010). *Wechsler Memory Scale- Fourth UK Edition.* San Antonio, TX: Pearson Assessment.

Wetter, M. W., & Corrigan, S. K. (1995). Providing information to clients about psychological tests: A survey of attorneys' and law students' attitudes. *Professional Psychology: Research and Practice, 26,* (5) 474-477.

Young, J. C., Sawyer, R. J., Roper, B. L., & Baughman, B. C. (2012). Expansion and re-examination of digit span effort indices on the WAIS-IV. *The Clinical Neuropsychologist, 26,* (1) 147-159.

Youngjohn, J. R., Lees-Haley, P. R., & Binder, L. M. (1999). Comment: Warning malingerers produces more sophisticated malingering. *Archives of Clinical Neuropsychology, 14,* 511-515.

Critical Appraisal

1. Defining the Field of Enquiry

I became interested in the controversies surrounding symptom validity testing during my first year clinical placement in a Traumatic Brain Injury (TBI) assessment service. I was intrigued by the idea that someone might fake or exaggerate their symptoms and by the fact that there was 'a test for that'. I recall feeling uncomfortable with the idea of covertly testing someone for effort/symptom validity. However, I was reassured by my supervisor who made it clear that someone's failure on a symptom validity test was not interpreted as evidence of malingering, rather that their test results might be invalid and need to be interpreted more cautiously.

During the first year on the Doctorate of Clinical Psychology course I was encouraged by course staff to generate ideas for suitable research projects. I was keen to embark on an ambitious and exciting project and had a preference for experimental, quantitative methods and an interest in neuropsychology. Despite these criteria, I found it difficult to settle on a single research idea, finding myself overwhelmed by too many possibilities. It was only through discussion with Professor Mike Wang that Symptom Validity Testing became an area of consideration. Professor Wang raised his concerns regarding the ethics of testing clients for symptom validity without informing them but suggested that to inform them may invalidate the test.

Exploration of the literature turned up very little previous research into the effects of informing consent. I came up with a simple experimental design, where I could ask "does informing clients that they will be tested for effort affect their symptom validity test results" by randomly assigning neuropsychology patients to either informed or uninformed conditions. Another, secondary question naturally evolved: Does

informing clients that they will be assessed for effort affect their memory test results? These questions seemed highly clinically relevant and useful to ensure both ethical and accurate assessment. I reasoned that if there was no difference between groups it could be argued that informing made no difference to test performance and should therefore be considered best practice. However, if informing clients that they would be tested for effort did affect their symptom validity test results further research could examine whether this was because the warning caused them to respond more honestly or malinger more effectively.

2. The Literature Review

In my formulation of a project proposal I soon found myself seeking definitions of malingering and base rate estimates. I was alarmed by several research papers using SVT test results alone to estimate base rates within a sample. This feeling was compounded by my further examination of SVTs and discovery that the cut-off scores for 'failure' were not below chance responding, but rather, below the floor performance (e.g. 45/50) of a normative sample. Most SVTs are tests of recognition memory, albeit very easy, if someone had a genuine inability to attend to, sense or encode the test stimuli and engaged in random responding they would, by chance, fail a test like the Test of Memory Malingering 90% of the time. While it might be unusual for a clinician to label such a patient as a 'malingerer', large scale studies, looking at test failure rates in isolation from other information might fall into this trap. Indeed, if base rate estimates within a given population are high, this may also have an impact on the clinician's willingness to diagnose malingering.

In my review of the literature I found an article entitled "have we forgotten the base rate problem?" (Rosenfeld, Sands & Van Gorp, 2000) I was both amused and

shocked by this articulation of a concern I was having twelve years on. I came to understand the issues around test sensitivity, specificity and positive predictive accuracy causing me to doubt the validity of SVTs altogether. Preliminary searches for studies mentioning 'base rates' and 'malingering' in their abstract revealed very few articles and I became aware that I would need to re-think my strategy. I found several articles referring to the Slick, Sherman and Iverson (1999) criteria for malingering and felt this was a fairly robust tool, with definite malingering only diagnosed when examinees score significantly below chance, have substantial external incentives to malinger, unusual or implausible test results and where all of the later cannot be explained by their neurological or psychiatric condition.

I embarked on a search for any studies that had either applied the Slick, Sherman and Iverson (1999) criteria or had provided enough information about their sample to allow its application. This was not as easy as I had anticipated, many of the studies providing information about base rates of malingering within a sampled population had done so in order to form a 'known groups design' for the validation of a symptom validity test. As such, search terms such as 'base rates' did not necessarily identify all of the relevant studies and I had to use overly inclusive terms in my database searching, reading all the returned abstracts, and often entire papers, to determine whether or not they contained base rate information. In between clinical placement and other academic responsibilities this process took months. However, once I had identified the useful research papers the write-up and analysis of them was relatively easy.

3. The Research Proposal

The chosen research question naturally dictated an experimental methodology. This was an approach to research that I felt fairly accustomed to and confident with having run a relatively large scale experiment previously for the purposes of my undergraduate degree. However, the practicalities associated with the use of a clinical sample were new to me. Asking other people to collect data for me, on the promise of authorship should the study be published felt like quite an imposition. I was concerned that asking clinicians to engage in this research was asking too much of them and would take control over data collection away from me. However, discussions with several Clinical Neuropsychologists and my research supervisor were encouraging. I was reassured that most of the tests I would be asking them to use were ones they commonly included for assessment purposes and that they were regularly receiving appropriate referrals. I also anticipated that having the burden of data collection taken 'off my hands' would make the process much easier. I consulted with the collaborating clinicians in deciding which tests to use; we opted for tests they had available and that were commonly used within their services.

At peer review concerns were raised that, given the sampled client group had memory difficulties, I needed to make sure that standardised 'informed' and 'uninformed' instructions were attended to and encoded by participants. The design was altered from just asking the clinician to read standardised instructions to them reading the instructions and presenting them in large written print for the participant to read. To ensure the participant had encoded these instructions they were then asked to explain the instructions back to the clinician.

Peer reviewers suggested that, given the volume of data I would be collecting, I should think of further research questions that might be addressed within the thesis. This was not a difficult task given, by this stage I had read around the subject of symptom validity testing extensively.

At no stage of the peer review process were any concerns raised regarding recruitment. Both peer reviewers seemed to think four sites would be an adequate source for the required number of participants.

4. Gaining Ethics Committee and Research and Development Approval

Course staff and trainee cohorts from previous years had prepared me to expect gaining ethical approval to take time. I was concerned that not informing participants of the true nature of the study and having 50% of participants uninformed that they would be tested for effort was a significant deception that might not be 'passed' by the ethics committee.

The ethics committee responded to my proposal in a relatively timely manner and I attended the committee meeting. While I was asked to explain the study in detail they didn't appear to have any real concerns about my proposal. They asked me to make some changes to participant information sheets and approved these within their standard timeframe.

I was not prepared for the difficulties I encountered in seeking approval from the Research and Development (R & D) departments of each collaborating NHS trust. I submitted my application along with supporting documents as requested to each trust as soon as I had final approval from the Ethics committee. I understood it might take some time for approval to be granted so did not follow up these applications until a month had

passed when I emailed the respective departments to check how my applications were progressing. None of the R & D departments had started processing my application and one responded asking: "what does it look like?"

It took eight months of chasing R & D departments to gain clearance across all sites. This was a delay I was not prepared for; I had assumed that R&D clearance would take a month or two to obtain.

5. Data Collection

By the time R & D clearance had been obtained the project was running well behind schedule, I was anxious to get data collection started. I met with two of the four services before Christmas, supplying them with paper copies of the study materials I had previously supplied in electronic form. One of the two other services arranged for me to visit in January 2012 to explain the procedure, which I did, agreeing to visit again once a month to maintain contact and address any 'teething issues'. I visited this service on two occasions but found that they had not started collecting data on either date. We discussed their difficulties in committing to the research and I offered to help with testing if they could gain consent from participants. They felt this would be helpful and set up several dates when I would be available to test participants. They arranged two participants for me to see at their site but, despite regular prompting, were unable to supply me with any more. Another trainee from my cohort was also on placement with this team and she was able to obtain a further two for my sample.

Throughout the recruitment phase I was aware that I was increasing the workload of participating services which had very little incentive to engage in the research; this limited the extent to which I felt I could be assertive about data collection. I sent emails to the non-responding service once every few weeks and eventually

received a positive response from the clinician who was willing to help but busy with an existing workload. I managed to arrange a meeting in March 2012 and went to see them with comprehensive 'data collection packs', I also offered to spend time assessing their clients to aid data collection. Despite these efforts the service was only able to provide two participants for the present study.

These difficulties were contrasted by the helpfulness of the other two services, one of which obtained 11 participants for the study and the other, which only saw one client a month managed to recruit all but one of the clients they saw within the data collection period.

6. Analysis of Results

I understood that, given my sample size, I would not achieve statistical power. Despite this knowledge, I was still disappointed when I examined the descriptive statistics and realised there would be no difference between groups. This made me wonder how I was going to write a 'passable' thesis. I decided to run the analyses as planned, not because it was necessary for interpretation of the data, but because it would demonstrate my ability to run appropriate analyses for the purposes of examination.

7. Write up

I was able to write the literature review, Introduction and Method sections of this paper in the months prior to data collection and analysis. While these sections took time and energy, they were not particularly difficult to write. I found the writing of the results and discussion sections much harder; in order to demonstrate the utility of the study I wanted to make interpretations regarding the data. However, any interpretation

of the data would have, at best, lacked rigour and, at worst been misleading and irresponsible.

8. Critique of the research

The present study aimed to address clinically relevant and theoretically driven research questions. Due to time limitations and recruitment problems the study did not achieve adequate statistical power. However, the act of speaking to clinicians and talking about the research has proved clinically useful and provoked further ideas for future research. The research questions remain important and data collection will continue with the aim of publishing findings once statistical power is reached.

In response to data collection concerns the inclusion criteria were broadened to include participants presenting with acquired, congenital and developmental brain conditions in addition to those with traumatic brain injuries. This may have made the sample more representative of persons presenting for neuropsychological assessment in the UK but may have also decreased statistical power by introducing further heterogeneity in the test scores of the sample, making it easier for a small to medium sized effect of 'informing' to be lost in large between-participant differences. The inclusion of participants without external incentive to deceive may have led to decreased variability in Test of Memory Malingering (TOMM; Tombaugh, 1996) scores. Without an increase in the variability of TOMM scores it will be difficult to assess which factors, if any, contribute to TOMM scores. With a larger sample more variability may become apparent, however, if it does not, sampling procedures may need reviewing with a view to sampling clients more at risk of malingering.

Aside from increasing the sample size, if I were to make any changes to the present study I would ask clinicians to provide participants scores for each item of the

Logical Memory-Recognition test (Wechsler, 2010). This would allow the application of the Rarely Missed Index (Killgore & Dellapietra, 2000) to the sample as an additional embedded measure of symptom validity. Including multiple embedded measures of symptom validity would allow analysis of the extent to which they are related to each other. If they are highly related it would suggest they are measuring the same construct and therefore not particularly useful as additional tests. If, however, they are not highly related, it would suggest they may measure different constructs. Such constructs would be worthy of further investigation in order to 'unpick' what symptom validity test failure actually means.

A further alteration to the present study would be to ask clinicians to determine whether any participant's symptom validity test failure could be explained by the clients existing psychological or neurological condition. If this information were collated it would satisfy the Slick Sherman and Iverson (1999) Criterion D for diagnosing definite and probable malingering. In the present data set one participant met: criterion 'A' by pursuing litigation; criterion 'B2' by scoring below cut offs on two symptom validity tests (and performing poorly on another). If we had collected information regarding criterion 'D' it would have been possible to determine whether this participant met the criteria for probable malingering. In the absence of UK base rates, any study providing a rigorous estimate of base rates would make a significant contribution to the field and our understanding of the positive predictive accuracy of symptom validity tests.

9. Learning Points

I knew from the start of this project that it was ambitious; however, I did not appreciate how ambitious it was. Had I have known how long R & D clearance would take and how difficult recruitment was going to be I would have opted for a 'safer'

project. A safer project would have been one where I was not dependant on other people to collect data for me, and where data collection could have been more quickly expedited. I could have investigated how clinical neuropsychologists deal with the conflicting advice from SVT manuals and BPS guidance by survey or interview methods. I could have asked clinical neuropsychologists how they interpret SVT test failure. I could have run the same experiment using student simulators, this would have been a less rigorous methodology, and more testing on my part but an easier sample to obtain.

I had no way of knowing how long R & D clearance would take, nor how recruitment would be a problem. With hindsight I realise that this project would be better undertaken by a, or even a group of, full-time neuropsychologists with access to a steady stream of potential participants and an unlimited time-frame for completion.

Despite my ideas for other projects I remain certain that the present project is methodologically sound and, given more time and resources, would produce interesting results worthy of publication. Whilst the research process has not always been easy I have enjoyed immersing myself in one area of the literature and getting to know it well. I have learned that, given time and enthusiasm, I could continue to make contributions to the literature post-qualifying. I have coped with and juggled the demands of clinical work, research and life in general. This has taught me a great deal about my own efficacy and ability to cope under pressure which I will take with me into future clinical roles.

10. References

British Psychological Society; Professional Practice Board (2009). *Assessment of Effort in Clinical Testing of Cognitive Functioning for Adults.* Leicester, UK: British Psychological Society.

Killgore, W. I. S., & Dellapietra, L. (2000). Using the WMS-III to detect malingering: Empirical validation of the Rarely Missed Index (RMI). *Journal of Clinical and Experimental Neuropsychology, 22,* (6), 761-771.

Rosenfeld, B., Sands, S. A., & Van Gorp, W.G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology, 15,* (4) 349-359.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13,* 545-561.

Tombaugh, T. N. (1996). *Test of Memory; TOMM.* North Tonawanda, NY: Multi-Health Systems Inc.

Wechsler, D. (2010). *Wechsler Memory Scale - Fourth UK Edition.* San Antonio, TX: Pearson Assessment.

Appendices

Appendix A : Participant Information Sheet

To be printed on trust headed paper)
Title: Factors Affecting Cognitive Test Results in Patients with Acquired Brain Injuries.
Participant Information Sheet (Version 2)
Last updated: 07/04/2011

Participant Information Sheet

Title of Research: Factors Affecting Cognitive Test Results in Patients with Acquired Brain Injuries.

Chief Investigator: Ms Alice Nicholls BSc (Hons)

We would like to invite you to take part in our research study. Before you decide we would like you to understand why the research is being done and what it would involve for you. One of our team will go through this sheet with you and answer any questions you have. This should take about 10 minutes.

What is the purpose of the study?

The term "cognitive functioning" refers to all the things your brain does. This includes your ability to direct your attention, perceive, learn, interpret and remember information. It also includes skills such as self control, planning and organising. We are often asked to assess the cognitive functioning of people who have had an acquired brain injury. It is important that these assessments are as accurate as possible. If the assessments are not accurate it may lead to someone not getting the treatment or support they need. Alternatively it may mean people get treatment that is inappropriate for them. Lots of different things may affect how accurate test results are. Our research aims to find out what some of these things are so we can continue to make the tests as accurate as possible. In order to do this the research will look at the relationship between your type of brain injury and your results on some of tests you will be taking.

What would be expected of me?

If you choose to take part in our research your treatment will remain the same. You will still be assessed and this assessment will be the same as you would have received if you had not chosen to take part. The only different thing we would ask you to do is agree to have your test results and information about your brain injury used for research purposes. The research is being conducted by a researcher at the Clinical Psychology Unit at the University of Leicester. This means that you would be allowing the researcher to access the details of your injury and your test results. The researcher will continue to keep any identifiable information about you confidential. If you decide to take part we will ask you to sign a consent form, this will be the only piece of information that identifies you. This will also go to the researcher at the University of Leicester but will be stored separately from your test results in a locked filing cabinet.

Do I have to take part?

No. If you choose not to take part it will not make any difference to your treatment.

If you choose to take part you can change your mind in the next 12 months by contacting your Neuropsychologist or the researcher and asking them to withdraw your data from the study. If you ask to withdraw your data, you will not have to give a reason and it will not affect the way you are treated.

Where can I get more information?

If you would like more information you can contact your Clinical Neuropsychologist or ask them to arrange for the researcher at the University of Leicester to contact you. You can contact the Researcher directly by emailing an154@le.ac.uk.

This research is being supervised by Prof. Wang, you can contact him by emailing mw125@le.ac.uk or by telephoning the Clinical Psychology Unit at the University of Leicester on 01162 231 639

Who has reviewed this study?

All research in the NHS is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by the Nottingham 1 Research Ethics Committee.

What should I do if I want to make a complaint?

If you have a complaint about the research study you can discuss it with your Clinical Neuropsychologist or the Chief investigator. You can contact the Chief investigator by leaving your contact details with your Clinical Neuropsychologist. If you remain unhappy and wish to make a formal complaint you can find ways of doing this by visiting the following website: www.nhs.uk/choiceintheNHS/Rightsandpledges/complaints.

What will happen to the results of the study?

The results of the study will be written up and submitted for publication in a peer reviewed journal. It will also be used by the researcher at the University of Leicester as a doctoral thesis and a copy will be kept by the University of Leicester. If you would like a summary of the study's findings to be sent to you please write your postal address on the consent form.

If you would like to take part in this research study please sign the consent form.

Appendix B: Participant Consent Form

(Form to be printed on headed paper)
Title: Factors Affecting Cognitive Test Results in Patients with Acquired Brain Injuries
Consent form (Version 2)
Centre Number:
Study Number:
(Last updated 18/04/2011)

Consent Form

Patient Identification Number for this trial:

Title of Project: Factors Affecting Cognitive Test Results in Patients with Acquired Brain Injuries

Name of Researcher: Ms Alice Nicholls BSc (Hons)

Please initial box

1. I confirm that I have read and understand the information sheet dated 07/04/2011 (version 2) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that relevant sections of my medical notes and data collected during the study, may be looked at by individuals from the University of Leicester, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐

4. I agree to take part in the above study. ☐

Name of Patient..... Date..... Signature.....
Name of Person
taking consent .....Date..... Signature.....

5. I would like to receive a summary of the study's findings ☐

If you would like to receive a summary of the study's findings please write your postal address in the space below:

When completed: 1 for participant; 1 for researcher site file; 1 (original) to be kept in medical notes.

Appendix C: Participant Debrief Sheet

(Form to be printed on headed paper)
Title: Factors Affecting Cognitive Test Results in Patients with Acquired Brain Injuries
Participant Debrief Sheet (Version 1)
(Last updated 19/01/2011)

Participant Debrief Sheet

Thank you for taking part in our research. Now that you have completed your assessment we are able to give you some more information about the research.

One of the tests you took today was an "effort test". Effort tests are routinely used by Clinical Neuropsychologists to see how hard people are trying during testing. This is important because if people do not try their hardest on all of the tests their assessment will not be accurate. This means that effort tests results are often used to judge whether a Neuropsychological assessment is accurate.

We are concerned that effort test results might be affected by factors other than how hard someone tries. One of these factors is the information given to clients about testing in the assessment. For example, the British Psychological Society recommends that all clients are informed that they will be assessed to see how hard they are trying. This seems to be good advice; however, the people who design the effort tests advise that the client should not be informed. This means that when Clinical Neuropsychologists inform their clients they may be making the effort test results inaccurate. If the effort test results are not accurate then it is not possible to tell whether the rest of the tests are accurate.

This research aims to find out whether informing clients that they will be assessed to see how hard they are trying makes any difference to their effort test results. To do this you will have either been told a) That you should try your hardest throughout testing or b) That you will be assessed to see how hard you are trying. We will pool your results with the results of approximately 50 other participants and see if there is a difference in effort test scores between people who were informed and people who were not informed. We will also examine whether the information affected any other test results. We will use the data about your injury to make sure that the two groups we are comparing are similar in the extent of their injuries. We are also interested to see if levels of anxiety and depression have had an additional effect on effort and cognitive test results. We will examine this by using your scores on a scale that identifies levels of anxiety and depression.

If you have any further questions about the research please speak to your Clinical Neuropsychologist. If you would like the Chief Investigator to telephone you please ask your Clinical Neuropsychologist to pass on your 'phone number to the Chief Investigator.

If you no longer wish to take part in the study please inform your Clinical Neuropsychologist who will remove your results from the study.

Thank you for taking part in the study.

Appendix D: Data Collection Form

NB: The collaborating Clinical Neuropsychologist is welcome to omit this sheet and send the researcher an anonymised (participant numbered) report.

<table>
<tr><td>Participant Number</td><td colspan="5"></td></tr>
<tr><td>Condition</td><td colspan="2">Informed</td><td colspan="3">Uninformed</td></tr>
<tr><td>Age</td><td colspan="5"></td></tr>
<tr><td>Sex</td><td colspan="5"></td></tr>
<tr><td>Years in Education</td><td colspan="5"></td></tr>
<tr><td>TOPF</td><td colspan="2">Raw Score:</td><td colspan="3">Standard Score:</td></tr>
<tr><td>If you did not use the TOPF please state which test you used and why</td><td colspan="5"></td></tr>
<tr><td>Type of Brain Injury</td><td colspan="5"></td></tr>
<tr><td>Date of Brain Injury</td><td colspan="5"></td></tr>
<tr><td>Length of PTA If applicable</td><td colspan="5"></td></tr>
<tr><td>Type of Neuroimaging</td><td colspan="2"></td><td colspan="3">How long after injury was this taken? (Please comment on most recent)</td></tr>
<tr><td rowspan="2">Neuroimaging shows evidence of which of the following? (Please circle as appropriate)</td><td>Focal Damage</td><td rowspan="2">Damage to Frontal Lobe</td><td rowspan="2">Damage to Parietal Lobe</td><td rowspan="2">Damage to Temporal Lobe</td><td rowspan="2">Damage to Occipital Lobe</td></tr>
<tr><td>General Damage</td></tr>
<tr><td>TOMM</td><td>Immediate Retention:</td><td></td><td>Delayed Retention:</td><td colspan="2"></td></tr>
<tr><td>WMS-IV Logical Memory I Recall Total Score</td><td>Raw Score:</td><td></td><td>Age Scaled Score:</td><td colspan="2"></td></tr>
<tr><td>WMS-IV Logical Memory II Recall Total Score</td><td>Raw Score:</td><td></td><td>Age Scaled Score:</td><td colspan="2"></td></tr>
<tr><td>WMS-IV Logical Memory Recognition Total Score</td><td>Raw Score:</td><td colspan="4"></td></tr>
</table>

WAIS-IV Digit Span Forward Total Score	Raw Score:			
WAIS-IV Digit Span Backwards	Raw Score:			
WAIS-IV Digit Span Sequencing	Raw Score:			
Digit Span Total	Raw Score:		Age Scaled Score:	
HADS	Score for Anxiety:		Score for Depression:	
State Anxiety at the start of the session:	/5	State Anxiety Immediately before the TOMM was administered:	/5	
State Anxiety at the End of the testing session:	/5			
BVRT	Number Correct Score:		Error Score:	
	Types of error	Frequency		
	Omissions:			
	Rotations:			
	Distortions:			
	Perseverations:			
	Misplacements:			
	Size:			
Is the Participant currently involved in any litigation regarding their brain injury?	Yes	No		
Has the client previously claimed compensation for this, or any other injury?				

Appendix E: Correspondences to and from the Ethics Committee

NHS
National Research Ethics Service

Telephone:
Facsimile:

17 March 2011

Trainee Clinical Psychologist
Leicestershire Partnership NHS Trust
104 Regent Road
Leicester
Leicestershire
LE1 7LT

Dear ..

Study Title: **Effort Test Results: The Effect of Informed Consent on a Brain Injured Sample**

REC reference number:

The Research Ethics Committee reviewed the above application at the meeting held on 08 March 2011. Thank you for attending to discuss the study.

Documents reviewed

The documents reviewed at the meeting were:

Document	*Version*	*Date*
Participant Information Sheet	1	19 January 2011
Investigator CV	..	12 February 2011
Investigator CV	-	12 February 2011
REC application	71195/189388/1/120	10 February 2011
Participant Debrief Sheet	1	19 January 2011
Protocol	4	12 February 2011
Participant Consent Form	1	19 January 2011

Discussion

- The Committee asked you to clarify how many participants are required for the study i.e. 50 or 80? You confirmed that 50 participants are required for statistical power. However, 80 would be ideal.
- The Committee informed you that they were concerned over data storage kept at your home. You confirmed that the data will be coded. The Committee agreed that it would be satisfactory for data to be held on the University server which is password protected etc.

- The Committee asked whether anything that had been raised at peer review, had been addressed in the application. You stated that you could provide written evidence of peer review should it be requested. However, you stated that there were no changes that needed to be addressed with the study methodology.

- The Committee informed you that the Consent Forms should be held for 7 years. However, raw data should be kept for the length the University requests. You questioned why Consent Forms should be kept for 7 years. The Committee stated that this would be included in any checks made by regulatory authorities for research governance purposes.

- The Committee asked for further information on the Memory Mallingering test e.g. if a patients does quite well, may there be a fear that something will be told to others? You stated that confidentiality will be adhered to and the test will be undertaken routinely. The Committee asked whether the results will be held in the clinical notes. You stated that the results will go in the medical notes as part of routine. Also, you confirmed that the results of the study will not go in the medical notes.

- The Committee informed you that you should explain what 'Cognitive functioning' is, in simple terms, in the Participant Information Sheet.

Provisional opinion

The Committee would be content to give a favourable ethical opinion of the research, subject to receiving a complete response to the request for further information set out below.

The Committee delegated authority to confirm its final opinion on the application to the Chair.

Further information or clarification required

1. The word 'you' is missing from the 2nd line of the 1st paragraph in the Participant Information Sheet to read 'Before you decide we would like you to understand why the research is being done....' This should be revised.

2. The term 'cognitive functioning' should be clearly defined in simple terms in the Participant Information Sheet.

When submitting your response to the Committee, please send revised documentation where appropriate underlining or otherwise highlighting the changes you have made and giving revised version numbers and dates.

If the committee has asked for clarification or changes to any answers given in the application form, please do not submit a revised copy of the application form; these can be addressed in a covering letter to the REC.

The Committee will confirm the final ethical opinion within a maximum of 60 days from the date of initial receipt of the application, excluding the time taken by you to respond fully to the above points. A response should be submitted by no later than 15 July 2011.

Membership of the Committee

The members of the Committee who were present at the meeting are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

11/EM/0049	**Please quote this number on all correspondence**

Yours sincerely

pp

Chair

Email:

Enclosures: *List of names and professions of members who were present at the meeting and those who submitted written comments.*

Copy to: *R&D Department for NHS care organisation at lead site –*

****

Trainee Clinical Psychologist

Clinical Psychology Unit

104 Regent Road

Leicester

LE1 7LT

7th of April 2011

Dear **********,

Study Title: **Effort Test Results: The Effect of Informed Consent on a Brain Injured Sample.**

REC Reference number: ****/**/******

Thank you for your response to the above application. I have enclosed a copy of the amended Participant Information Sheet as requested.

If you have any further queries please feel free to contact me.

Yours Sincerely

****** ********

Trainee Clinical Psychologist

NHS

National Research Ethics Service

NRES Committee

Telephone: (Direct Line)
Facsimile:

20 April 2011

Trainee Clinical Psychologist
Leicester Partnership NHS Trust
104 Regent Road
Leicester
Leicestershire
LE1 7LT

Dear

Study title: **Effort Test Results: The Effect of Informed Consent on a Brain Injured Sample**

REC reference:

Thank you for your letter of 07 April 2011, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites

NHS sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS sites

The Committee has not yet been notified of the outcome of any site-specific assessment (SSA) for the non-NHS research site(s) taking part in this study. The favourable opinion does not therefore apply to any non-NHS site at present. We will write to you again as soon as one Research Ethics Committee has notified the outcome of a SSA. In the meantime no study procedures should be initiated at non-NHS sites.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

This Research Ethics Committee is an advisory committee tc uthority

The National Research Ethics Service (NRES) represents the NRES Directorate within the National Patient Safety Agency and Research Ethics Committees in England

WPH 1370

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Protocol	4	12 February 2011
Response to Request for Further Information		07 April 2011
REC application		10 February 2011
Participant Debrief Sheet	1	19 January 2011
Participant Information Sheet	2	07 April 2011
Investigator CV		12 February 2011
Investigator CV		12 February 2011
Participant Consent Form	2	18 April 2011

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@ nhs.uk.

Please quote this number on all correspondence

With the Committee's best wishes for the success of this project

Yours sincerely

pp

Chair

Email: nhs.uk

Enclosures: "After ethical review – guidance for researchers"

Copy to: – *Academic Supervisor*

R&D office for NHS care organisation at lead site –

Appendix F: Primary Hypothesis Calculations

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
TOMM1	.401	17	.000	.596	17	.000
TOMM2	.518	17	.000	.399	17	.000
LM1SS	.177	17	.160	.911	17	.103
LM2SS	.155	17	.200*	.920	17	.150
LMRTS	.167	17	.200*	.892	17	.050
DSSS	.113	17	.200*	.979	17	.952
RDSR	.098	17	.200*	.985	17	.991
HADSA	.105	17	.200*	.985	17	.988
HADSD	.155	17	.200*	.966	17	.750
BVRTcorrect	.147	17	.200*	.956	17	.563
BVRTerror	.194	17	.087	.905	17	.083
Omissions	.442	17	.000	.540	17	.000
Distortions	.259	17	.004	.775	17	.001
Rotations	.247	17	.007	.838	17	.007
Perseveration	.462	17	.000	.548	17	.000
Misplacements	.345	17	.000	.478	17	.000
Size	.537	17	.000	.262	17	.000

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Nonparametric Correlations between Variables

Correlations

		Age	TOMM1	TOMM2	Education	PMIQ	monthsinceInj	PTAlength	LM1SS	LM2SS	LMRTS	DSSS	HADSA	HADSD	BVRTcorrect	BVRTError
Kendall's tau_b	Age	Correlation Coefficient	1.000	.000	.047	-.035	.093	-.026	.042	.207	.171	.298	-.075	.055	-.074	.169	-.143

Correlations

		Age	TOMM1	TOMM2	Education	PMIQ	monthsinceInj	PTALength	LM1SS	LM2SS	LMRTS	DSSS	HADSA	HADSD	BVRTcorrect	BVRTerror
	Sig. (2-tailed)	.	1.000	.802	.843	.577	.891	.813	.209	.298	.071	.647	.743	.667	.375	.449
	N	21	21	21	21	20	16	21	21	21	21	21	20	20	17	17
TOMM1	Correlation Coefficient	.000	1.000	.602**	.266	.203	-.050	-.066	.474**	.590**	.312	.180	-.084	-.050	.206	-.215
	Sig. (2-tailed)	1.000	.	.003	.162	.264	.803	.728	.008	.001	.080	.311	.644	.786	.321	.297
	N	21	21	21	21	20	16	21	21	21	21	21	20	20	17	17
TOMM2	Correlation Coefficient	.047	.602**	1.000	.142	.382*	-.272	.000	.473*	.527**	.360	.338	-.122	.287	.294	-.290
	Sig. (2-tailed)	.802	.003	.	.481	.045	.205	1.000	.012	.005	.056	.070	.528	.144	.175	.176
	N	21	21	21	21	20	16	21	21	21	21	21	20	20	17	17
Education	Correlation Coefficient	-.035	.266	.142	1.000	.579**	.097	.103	.282	.245	.094	.128	.203	.274	.281	-.222

Correlations

		Age	TOMM1	TOMM2	Education	PMIQ	monthsinceInj	PTAlength	LM1SS	LM2SS	LMRTS	DSSS	HADSA	HADSD	BVRTcorrect	BVRTerror
	Sig. (2-tailed)	.843	.162	.481	.	.001	.627	.583	.112	.164	.596	.467	.264	.137	.169	.273
	N	21	21	21	21	20	16	21	21	21	21	21	20	20	17	17
PMIQ	Correlation Coefficient	.093	.203	.382*	.579**	1.000	-.137	.114	.504**	.363*	.356*	.413*	.012	.259	.372*	-.313
	Sig. (2-tailed)	.577	.264	.045	.001	.	.485	.529	.003	.031	.035	.014	.944	.136	.048	.094
	N	20	20	20	20	20	15	20	20	20	20	20	19	19	17	17
monthsinceInj	Correlation Coefficient	-.026	-.050	-.272	.097	-.137	1.000	-.205	-.201	-.087	-.026	-.251	.234	.170	-.066	.173
	Sig. (2-tailed)	.891	.803	.205	.627	.485	.	.298	.293	.648	.891	.187	.232	.392	.778	.445
	N	16	16	16	16	15	16	16	16	16	16	16	15	15	12	12
PTAlength	Correlation Coefficient	.042	-.066	.000	.103	.114	-.205	1.000	-.018	-.030	-.144	.226	-.034	.117	.270	-.257

Correlations

		Age	TOMM1	TOMM2	Education	PMIQ	monthsinceInj	PTAlength	LM1SS	LM2SS	LMRTS	DSSS	HADSA	HADSD	BVRTcorrect	BVRTerror
	Sig. (2-tailed)	.813	.728	1.000	.583	.529	.298	.	.919	.865	.415	.199	.853	.526	.186	.204
	N	21	21	21	21	20	16	21	21	21	21	21	20	20	17	17
LM1SS	Correlation Coefficient	.207	.474**	.473*	.282	.504**	-.201	-.018	1.000	.812**	.658**	.415*	.168	.292	.259	-.232
	Sig. (2-tailed)	.209	.008	.012	.112	.003	.293	.919	.	.000	.000	.012	.323	.091	.176	.221
	N	21	21	21	21	20	16	21	21	21	21	21	20	20	17	17
LM2SS	Correlation Coefficient	.171	.590**	.527**	.245	.363*	-.087	-.030	.812**	1.000	.645**	.388*	.100	.227	.257	-.183
	Sig. (2-tailed)	.298	.001	.005	.164	.031	.648	.865	.000	.	.000	.019	.554	.185	.177	.333
	N	21	21	21	21	20	16	21	21	21	21	21	20	20	17	17
LMRTS	Correlation Coefficient	.298	.312	.360	.094	.356*	-.026	-.144	.658**	.645**	1.000	.299	.101	.218	.082	-.040

Correlations

		Age	TOMM1	TOMM2	Education	PMIQ	monthsinceInj	PTALength	LM1SS	LM2SS	LMRTS	DSSS	HADSA	HADSD	BVRTcorrect	BVRTerror
	Sig. (2-tailed)	.071	.080	.056	.596	.035	.891	.415	.000	.000	.	.071	.553	.207	.671	.832
	N	21	21	21	21	20	16	21	21	21	21	21	20	20	17	17
DSSS	Correlation Coefficient	-.075	.180	.338	.128	.413*	-.251	.226	.415*	.388*	.299	1.000	-.100	.227	.073	-.040
	Sig. (2-tailed)	.647	.311	.070	.467	.014	.187	.199	.012	.019	.071	.	.555	.186	.704	.833
	N	21	21	21	21	20	16	21	21	21	21	21	20	20	17	17
HADSA	Correlation Coefficient	.055	-.084	-.122	.203	.012	.234	-.034	.168	.100	.101	-.100	1.000	.407*	.080	-.087
	Sig. (2-tailed)	.743	.644	.528	.264	.944	.232	.853	.323	.554	.553	.555	.	.017	.674	.644
	N	20	20	20	20	19	15	20	20	20	20	20	20	20	17	17
HADSD	Correlation Coefficient	-.074	-.050	.287	.274	.259	.170	.117	.292	.227	.218	.227	.407*	1.000	-.041	.146

Correlations

		Age	TOMM1	TOMM2	Education	PMIQ	monthsinceInj	PTALength	LM1SS	LM2SS	LMRTSS	DSSS	HADSA	HADSD	BVRTcorrect	BVRTerror
	Sig. (2-tailed)	.667	.786	.144	.137	.136	.392	.526	.091	.185	.207	.186	.017	.	.832	.446
	N	20	20	20	20	19	15	20	20	20	20	20	20	20	17	17
BVRTcorrect	Correlation Coefficient	.169	.206	.294	.281	.372*	-.066	.270	.259	.257	.082	.073	.080	-.041	1.000	-.841**
	Sig. (2-tailed)	.375	.321	.175	.169	.048	.778	.186	.176	.177	.671	.704	.674	.832	.	.000
	N	17	17	17	17	17	12	17	17	17	17	17	17	17	17	17
BVRTerror	Correlation Coefficient	-.143	-.215	-.290	-.222	-.313	.173	-.257	-.232	-.183	-.040	-.040	-.087	.146	-.841**	1.000
	Sig. (2-tailed)	.449	.297	.176	.273	.094	.445	.204	.221	.333	.832	.833	.644	.446	.000	.
	N	17	17	17	17	17	12	17	17	17	17	17	17	17	17	17

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of TOMM2 is the same across categories of Information.	Independent-Samples Mann-Whitney U Test	.726	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

General Linear Model

Between-Subjects Factors

		Value Label	N
Information	1	informed	8
	2	control	9

Descriptive Statistics

	Information	Mean	Std. Deviation	N
LM2SS	informed	7.8750	4.18970	8
	control	7.5556	3.84419	9
	Total	7.7059	3.88530	17
DSSS	informed	8.0000	3.85450	8
	control	9.3333	2.00000	9
	Total	8.7059	2.99509	17
BVRTcorrect	informed	7.2500	1.66905	8
	control	7.0000	2.23607	9
	Total	7.1176	1.93269	17

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Information

Box's Test of Equality of Covariance Matrices[a]

Box's M	10.841
F	1.406
df1	6
df2	1555.544
Sig.	.209

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Intercept	Pillai's Trace	.960	102.724[a]	3.000	13.000	.000	.960	308.171	1.000
	Wilks' Lambda	.040	102.724[a]	3.000	13.000	.000	.960	308.171	1.000
	Hotelling's Trace	23.705	102.724[a]	3.000	13.000	.000	.960	308.171	1.000
	Roy's Largest Root	23.705	102.724[a]	3.000	13.000	.000	.960	308.171	1.000
Information	Pillai's Trace	.079	.373[a]	3.000	13.000	.774	.079	1.118	.105
	Wilks' Lambda	.921	.373[a]	3.000	13.000	.774	.079	1.118	.105
	Hotelling's Trace	.086	.373[a]	3.000	13.000	.774	.079	1.118	.105
	Roy's Largest Root	.086	.373[a]	3.000	13.000	.774	.079	1.118	.105

a. Exact statistic

b. Computed using alpha = .05

c. Design: Intercept + Information

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	LM2SS	.432[a]	1	.432	.027	.872	.002	.027	.053
	DSSS	7.529[c]	1	7.529	.830	.377	.052	.830	.137
	BVRTcorrect	.265[d]	1	.265	.067	.800	.004	.067	.057
Intercept	LM2SS	1008.432	1	1008.432	62.740	.000	.807	62.740	1.000
	DSSS	1272.471	1	1272.471	140.346	.000	.903	140.346	1.000
	BVRTcorrect	860.029	1	860.029	216.814	.000	.935	216.814	1.000
Information	LM2SS	.432	1	.432	.027	.872	.002	.027	.053
	DSSS	7.529	1	7.529	.830	.377	.052	.830	.137
	BVRTcorrect	.265	1	.265	.067	.800	.004	.067	.057
Error	LM2SS	241.097	15	16.073					
	DSSS	136.000	15	9.067					
	BVRTcorrect	59.500	15	3.967					
Total	LM2SS	1251.000	17						
	DSSS	1432.000	17						
	BVRTcorrect	921.000	17						
Corrected Total	LM2SS	241.529	16						
	DSSS	143.529	16						
	BVRTcorrect	59.765	16						

a. R Squared = .002 (Adjusted R Squared = -.065)

b. Computed using alpha = .05

c. R Squared = .052 (Adjusted R Squared = -.011)

d. R Squared = .004 (Adjusted R Squared = -.062)

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
LM2SS	.049	1	15	.828
DSSS	3.311	1	15	.089
BVRTcorrect	.796	1	15	.386

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Information

Appendix G: Test of Group Equivalence

Correlations

			HADSA	HADSD	Age	Education	PMIQ	monthsincelnj	PTAlength
Kendall's tau_b	HADSA	Correlation Coefficient	1.000	.407*	.055	.203	.012	.234	-.034
		Sig. (2-tailed)	.	.017	.743	.264	.944	.232	.853
		N	20	20	20	20	19	15	20
	HADSD	Correlation Coefficient	.407*	1.000	-.074	.274	.259	.170	.117
		Sig. (2-tailed)	.017	.	.667	.137	.136	.392	.526
		N	20	20	20	20	19	15	20
	Age	Correlation Coefficient	.055	-.074	1.000	-.035	.093	-.026	.042
		Sig. (2-tailed)	.743	.667	.	.843	.577	.891	.813
		N	20	20	21	21	20	16	21
	Education	Correlation Coefficient	.203	.274	-.035	1.000	.579**	.097	.103
		Sig. (2-tailed)	.264	.137	.843	.	.001	.627	.583
		N	20	20	21	21	20	16	21
	PMIQ	Correlation Coefficient	.012	.259	.093	.579**	1.000	-.137	.114
		Sig. (2-tailed)	.944	.136	.577	.001	.	.485	.529
		N	19	19	20	20	20	15	20

	monthsinceInj	Correlation Coefficient	.234	.170	-.026	.097	-.137	1.000	-.205
		Sig. (2-tailed)	.232	.392	.891	.627	.485	.	.298
		N	15	15	16	16	15	16	16
	PTAlength	Correlation Coefficient	-.034	.117	.042	.103	.114	-.205	1.000
		Sig. (2-tailed)	.853	.526	.813	.583	.529	.298	.
		N	20	20	21	21	20	16	21

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Between-Subjects Factors

		Value Label	N
Information	1	informed	6
	2	control	8

Descriptive Statistics

	Information	Mean	Std. Deviation	N
Age	informed	50.83	6.014	6
	control	36.75	12.209	8
	Total	42.79	12.103	14
monthsinceInj	informed	15.6667	15.21403	6
	control	59.8750	88.66701	8
	Total	40.9286	69.55403	14
PTAlength	informed	13.5000	28.78020	6
	control	17.6250	41.67883	8
	Total	15.8571	35.47449	14
HADSD	informed	7.1667	5.11534	6
	control	11.0000	3.85450	8
	Total	9.3571	4.68397	14
PMIQ	informed	97.3333	7.58068	6

	control	97.5000	10.55597	8
	Total	97.4286	9.06145	14

Box's Test of Equality of Covariance Matrices[a]

Box's M	38.040
F	1.284
df1	15
df2	464.102
Sig.	.208

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Information

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Intercept	Pillai's Trace	.994	247.620[a]	5.000	8.000	.000	.994	1238.102	1.000
	Wilks' Lambda	.006	247.620[a]	5.000	8.000	.000	.994	1238.102	1.000
	Hotelling's Trace	154.763	247.620[a]	5.000	8.000	.000	.994	1238.102	1.000
	Roy's Largest Root	154.763	247.620[a]	5.000	8.000	.000	.994	1238.102	1.000
Information	Pillai's Trace	.624	2.659[a]	5.000	8.000	.105	.624	13.296	.510
	Wilks' Lambda	.376	2.659[a]	5.000	8.000	.105	.624	13.296	.510

	Hotelling's Trace	1.662	2.659[a]	5.000	8.000	.105	.624	13.296	.510
	Roy's Largest Root	1.662	2.659[a]	5.000	8.000	.105	.624	13.296	.510

a. Exact statistic

b. Computed using alpha = .05

c. Design: Intercept + Information

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Age	3.654	1	12	.080
monthsinceInj	2.463	1	12	.143
PTAlength	.174	1	12	.684
HADSD	.749	1	12	.404
PMIQ	.747	1	12	.404

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Information

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	Age	680.024[a]	1	680.024	6.665	.024	.357	6.665	.660
	monthsinceInj	6700.720[c]	1	6700.720	1.431	.255	.107	1.431	.197
	PTAlength	58.339[d]	1	58.339	.043	.839	.004	.043	.054
	HADSD	50.381[e]	1	50.381	2.574	.135	.177	2.574	.315

	PMIQ	.095[f]	1	.095	.001	.974	.000	.001	.050
Intercept	Age	26300.024	1	26300.024	257.773	.000	.956	257.773	1.000
	monthsinceInj	19565.292	1	19565.292	4.178	.064	.258	4.178	.468
	PTAlength	3321.482	1	3321.482	2.445	.144	.169	2.445	.302
	HADSD	1131.524	1	1131.524	57.821	.000	.828	57.821	1.000
	PMIQ	130148.667	1	130148.667	1463.258	.000	.992	1463.258	1.000
Information	Age	680.024	1	680.024	6.665	.024	.357	6.665	.660
	monthsinceInj	6700.720	1	6700.720	1.431	.255	.107	1.431	.197
	PTAlength	58.339	1	58.339	.043	.839	.004	.043	.054
	HADSD	50.381	1	50.381	2.574	.135	.177	2.574	.315
	PMIQ	.095	1	.095	.001	.974	.000	.001	.050
Error	Age	1224.333	12	102.028					
	monthsinceInj	56190.208	12	4682.517					
	PTAlength	16301.375	12	1358.448					
	HADSD	234.833	12	19.569					
	PMIQ	1067.333	12	88.944					
Total	Age	27533.000	14						
	monthsinceInj	86343.000	14						
	PTAlength	19880.000	14						
	HADSD	1511.000	14						
	PMIQ	133960.000	14						

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	Age	680.024[a]	1	680.024	6.665	.024	.357	6.665	.660
	monthsincelnj	6700.720[c]	1	6700.720	1.431	.255	.107	1.431	.197
	PTAlength	58.339[d]	1	58.339	.043	.839	.004	.043	.054
	HADSD	50.381[e]	1	50.381	2.574	.135	.177	2.574	.315
	PMIQ	.095[f]	1	.095	.001	.974	.000	.001	.050
Intercept	Age	26300.024	1	26300.024	257.773	.000	.956	257.773	1.000
	monthsincelnj	19565.292	1	19565.292	4.178	.064	.258	4.178	.468
	PTAlength	3321.482	1	3321.482	2.445	.144	.169	2.445	.302
	HADSD	1131.524	1	1131.524	57.821	.000	.828	57.821	1.000
	PMIQ	130148.667	1	130148.667	1463.258	.000	.992	1463.258	1.000
Information	Age	680.024	1	680.024	6.665	.024	.357	6.665	.660
	monthsincelnj	6700.720	1	6700.720	1.431	.255	.107	1.431	.197
	PTAlength	58.339	1	58.339	.043	.839	.004	.043	.054
	HADSD	50.381	1	50.381	2.574	.135	.177	2.574	.315
	PMIQ	.095	1	.095	.001	.974	.000	.001	.050
Corrected Total	Age	1904.357	13						
	monthsincelnj	62890.929	13						
	PTAlength	16359.714	13						

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	Age	680.024[a]	1	680.024	6.665	.024	.357	6.665	.660
	monthsincelnj	6700.720[c]	1	6700.720	1.431	.255	.107	1.431	.197
	PTAlength	58.339[d]	1	58.339	.043	.839	.004	.043	.054
	HADSD	50.381[e]	1	50.381	2.574	.135	.177	2.574	.315
	PMIQ	.095[f]	1	.095	.001	.974	.000	.001	.050
Intercept	Age	26300.024	1	26300.024	257.773	.000	.956	257.773	1.000
	monthsincelnj	19565.292	1	19565.292	4.178	.064	.258	4.178	.468
	PTAlength	3321.482	1	3321.482	2.445	.144	.169	2.445	.302
	HADSD	1131.524	1	1131.524	57.821	.000	.828	57.821	1.000
	PMIQ	130148.667	1	130148.667	1463.258	.000	.992	1463.258	1.000
Information	Age	680.024	1	680.024	6.665	.024	.357	6.665	.660
	monthsincelnj	6700.720	1	6700.720	1.431	.255	.107	1.431	.197
	PTAlength	58.339	1	58.339	.043	.839	.004	.043	.054
	HADSD	50.381	1	50.381	2.574	.135	.177	2.574	.315
	PMIQ	.095	1	.095	.001	.974	.000	.001	.050
	HADSD	285.214	13						
	PMIQ	1067.429	13						

a. R Squared = .357 (Adjusted R Squared = .304)

b. Computed using alpha = .05

c. R Squared = .107 (Adjusted R Squared = .032)

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	Age	680.024[a]	1	680.024	6.665	.024	.357	6.665	.660
	monthsincelnj	6700.720[c]	1	6700.720	1.431	.255	.107	1.431	.197
	PTAlength	58.339[d]	1	58.339	.043	.839	.004	.043	.054
	HADSD	50.381[e]	1	50.381	2.574	.135	.177	2.574	.315
	PMIQ	.095[f]	1	.095	.001	.974	.000	.001	.050
Intercept	Age	26300.024	1	26300.024	257.773	.000	.956	257.773	1.000
	monthsincelnj	19565.292	1	19565.292	4.178	.064	.258	4.178	.468
	PTAlength	3321.482	1	3321.482	2.445	.144	.169	2.445	.302
	HADSD	1131.524	1	1131.524	57.821	.000	.828	57.821	1.000
	PMIQ	130148.667	1	130148.667	1463.258	.000	.992	1463.258	1.000
Information	Age	680.024	1	680.024	6.665	.024	.357	6.665	.660
	monthsincelnj	6700.720	1	6700.720	1.431	.255	.107	1.431	.197
	PTAlength	58.339	1	58.339	.043	.839	.004	.043	.054
	HADSD	50.381	1	50.381	2.574	.135	.177	2.574	.315
	PMIQ	.095	1	.095	.001	.974	.000	.001	.050

d. R Squared = .004 (Adjusted R Squared = -.079)

e. R Squared = .177 (Adjusted R Squared = .108)

f. R Squared = .000 (Adjusted R Squared = -.083)

Appendix H: Raw Data for Litigating and Non-Litigating Participants

Case Summaries[a]

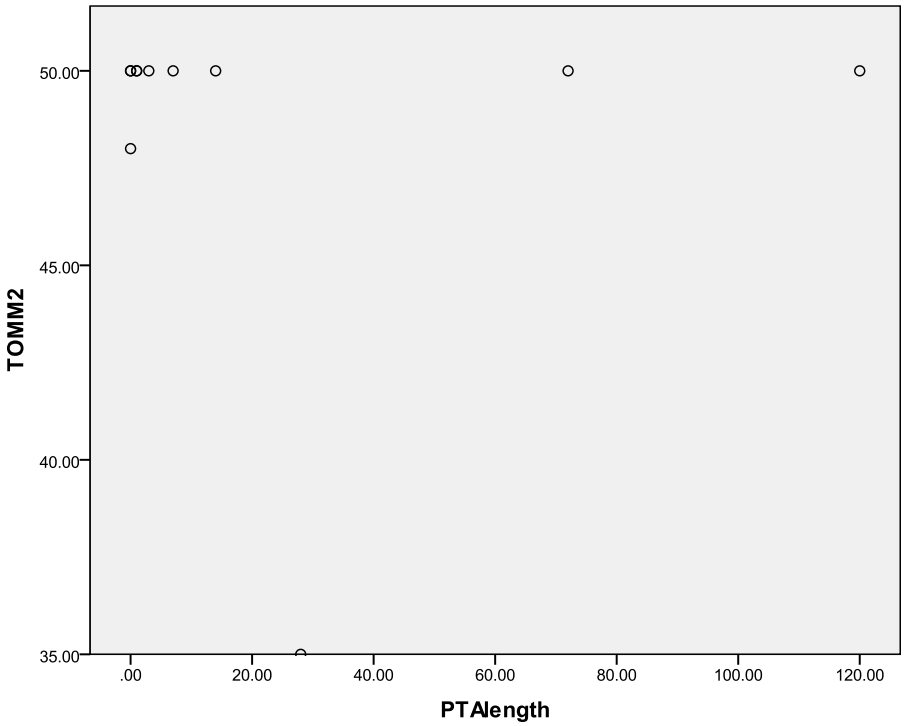
				LMRTS	RDSR	TOMM2	Omissions	Rotations	Distortions	Perseveration	Misplacements	Size	BVRT error
Litigating	yes	1		23.00	14.00	50.00	1.00	.00	4.00	.00	.00	.00	5.00
		2		29.00	22.00	50.00	.00	.00	.00	.00	1.00	.00	1.00
		3		29.00	14.00	50.00	.00	2.00	1.00	2.00	.00	.00	5.00
		4		25.00	18.00	50.00	.00	1.00	4.00	.00	.00	.00	5.00
		5		15.00	10.00	35.00	.	.	.	.	.	.	.
		6		23.00	17.00	50.00	.00	.00	.00	.00	.00	1.00	1.00
		7		30.00	25.00	50.00	.	.	.	.	.	.	.
		Total	N	7	7	7	5	5	5	5	5	5	5
			Mean	24.8571	17.1429	47.8571	.2000	.6000	1.8000	.4000	.2000	.2000	3.4000
			Std. Deviation	5.24177	5.11301	5.66947	.44721	.89443	2.04939	.89443	.44721	.44721	2.19089
	no	1		26.00	22.00	50.00	.	.	.	.	.	.	.
		2		28.00	19.00	50.00	2.00	2.00	.00	2.00	1.00	.00	7.00
		3		29.00	12.00	50.00	.00	.00	2.00	.00	.00	.00	2.00
		4		23.00	17.00	50.00	.00	.00	.00	.00	.00	.00	.00
		5		18.00	10.00	50.00	.00	.00	5.00	.00	1.00	.00	4.00
		6		23.00	15.00	47.00	.00	1.00	1.00	1.00	.00	.00	5.00
		7		17.00	18.00	50.00	.00	1.00	.00	2.00	1.00	.00	4.00

Case Summaries[a]

			LMRTS	RDSR	TOMM2	Omissions	Rotations	Distortions	Perseveration	Misplacements	Size	BVRT error
	8		19.00	16.00	48.00	5.00	2.00	.00	.00	1.00	.00	9.00
	9		26.00	13.00	50.00	4.00	1.00	3.00	.00	.00	.00	8.00
	10		29.00	15.00	50.00	.00	1.00	1.00	.00	.00	.00	2.00
	11		22.00	14.00	50.00	.	.	.	.	.	.	.
	12		27.00	16.00	50.00	.00	.00	.00	.00	.00	.00	.00
	13		17.00	15.00	50.00	.00	2.00	.00	.00	.00	.00	2.00
	14		24.00	21.00	50.00	.00	3.00	5.00	.00	6.00	.00	14.00
	Total	N	14	14	14	12	12	12	12	12	12	12
		Mean	23.4286	15.9286	49.6429	.9167	1.0833	1.4167	.4167	.8333	.0000	4.7500
		Std. Deviation	4.32727	3.31580	.92878	1.78164	.99620	1.92865	.79296	1.69670	.00000	4.13686
Total	N		21	21	21	17	17	17	17	17	17	17
	Mean		23.9048	16.3333	49.0476	.7059	.9412	1.5294	.4118	.6471	.0588	4.3529
	Std. Deviation		4.57061	3.91578	3.30872	1.53153	.96635	1.90780	.79521	1.45521	.24254	3.65618

a. Limited to first 100 cases.

Appendix I: Relationships between Severity of Injury and TOMM-II Scores



Correlations

			PTAlength	TOMM2
Kendall's tau_b	PTAlength	Correlation Coefficient	1.000	.032
		Sig. (2-tailed)	.	.906
		N	11	11
	TOMM2	Correlation Coefficient	.032	1.000
		Sig. (2-tailed)	.906	.
		N	11	11

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of TOMM2 is the same across categories of yesnoevidence.	Independent-Samples Mann-Whitney U Test	.244	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Appendix J: The Relationship between SVT Scores

Nonparametric Correlations

Correlations

			TOMM2	LMRTS	RDSR	BVRTerror
Kendall's tau_b	TOMM2	Correlation Coefficient	1.000	.360	.235	-.290
		Sig. (2-tailed)	.	.056	.209	.176
		N	21	21	21	17
	LMRTS	Correlation Coefficient	.360	1.000	.273	-.040
		Sig. (2-tailed)	.056	.	.098	.832
		N	21	21	21	17
	RDSR	Correlation Coefficient	.235	.273	1.000	-.008
		Sig. (2-tailed)	.209	.098	.	.967
		N	21	21	21	17
	BVRTerror	Correlation Coefficient	-.290	-.040	-.008	1.000
		Sig. (2-tailed)	.176	.832	.967	.
		N	17	17	17	17

Appendix K: Statement of Epistemological Position

The present research was conducted from a positivist epistemological position: It assumes that symptom validity and brain functioning can be meaningfully measured. Furthermore it assumes that if informing individuals they will be assessed for effort affects their symptom validity test results such an effect would be measurable. In keeping with the positivist tradition this research has sought to support or falsify hypotheses through the quantitative analysis of validated measures.

Appendix L: Research Process Chronology

Research Proposal	**June 2010**
Peer Review	**July 2010**
Ethics Application submission	**February 2011**
Ethics committee meeting	**March 2011**
Responding to requests from ethics committee	**April 2011**
NHS Ethics approval received	**April 2011**
R&D application submission	**April 2011**
Literature review database searches	**October-December**
R&D approval received	**December 2012**
Data collection	**December-June**
Writing up of literature review	**January- March**
Writing up of research paper	**March – July 2012**

Appendix M: Guidelines to Authors for Journal Targeted for Literature Review

Neuropsychology Review

Editor-in-Chief: Edith V. Sullivan

ISSN: 1040-7308 (print version)
ISSN: 1573-6660 (electronic version)

Instructions for Authors

TYPES OF PAPERS

Review, Editorial, Commentary, Preface

SINGLE -BLIND PEER REVIEW

This journal follows a single-blind reviewing procedure. Authors are therefore requested to submit a title page, containing title, all author names, affiliations, and the contact information of the corresponding author. Any acknowledgements, disclosures, or funding information should also be included on this page.

MANUSCRIPT SUBMISSION

Manuscript Submission

Submission of a manuscript implies: that the work described has not been published before; that it is not under consideration for publication anywhere else; that its publication has been approved by all co-authors, if any, as well as by the responsible authorities – tacitly or explicitly – at the institute where the work has been carried out. The publisher will not be held legally responsible should there be any claims for compensation.

Permissions

Authors wishing to include figures, tables, or text passages that have already been published elsewhere are required to obtain permission from the copyright owner(s) for both the print and online format and to include evidence that such permission has been granted when submitting their papers. Any material received without such evidence will be assumed to originate from the authors.

Online Submission

Authors should submit their manuscripts online. Electronic submission substantially reduces the editorial processing and reviewing times and shortens overall publication times. Please follow the hyperlink “Submit online” on the right and upload all of your manuscript files following the instructions given on the screen.

TITLE PAGE

Title Page

The title page should include:

- The name(s) of the author(s)
- A concise and informative title
- The affiliation(s) and address(es) of the author(s)
- The e-mail address, telephone and fax numbers of the corresponding author

Abstract

Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references.
Please provide 4 to 6 keywords which can be used for indexing purposes.