MALINGERING OF HEAD INJURY ON NEUROPSYCHOLOGICAL INSTRUMENTS: A META-ANALYTIC REVIEW

by

Lynne Elaine Sullivan

B.A., University of Manitoba, 1994

M.A., Simon Fraser University, 1996

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in the Department

of

Psychology

SIMON FRASER UNIVERSITY

October, 2000

ABSTRACT

There is a relatively large body of literature investigating malingering of head injury using neuropsychological instruments (Faust & Ackley, 1998; Nies & Sweet, 1994; Rogers, Harrell, & Liff, 1993). To date, there has been no consensus in the literature as to which instrument(s) is superior at distinguishing between head injury malingerers and non-malingerers. Furthermore, although a number of reviews have addressed the effectiveness of different strategies for identifying malingering, no meta-analytic studies of the literature have been undertaken to identify which measures perform best at detecting malingerers. The purpose of this study was, first of all, to identify such measures. Malingerers were compared both to normal controls and to brain-injured individuals. Second, the effectiveness of instruments from different cognitive domains in identifying malingerers was investigated along with the interaction between type of instrument and type of malingerer (Naïve, Coached, or Suspected malingerers, and Litigants). The moderating influence of study and participant characteristics was also investigated. Results indicated that malingerers could best be distinguished from non-malingerers by recognition tasks (e.g., the Recognition Memory Test), as well as tests designed specifically to assess malingering (e.g., the Portland Digit Recognition Test). On malingering tests, the differences between types of malingerers were eliminated, but on other neuropsychological instruments, Coached participants performed significantly worse than did other types of malingering participants. Also on malingering tests, there were no differences among malingering groups when compared to either normal or brain-injured comparison groups. When malingerers were compared to normals on

other neuropsychological instruments, effect sizes were larger than when malingerers were compared to brain-injured individuals. Overall, the results suggest that at least some neuropsychological instruments are valid for distinguishing between malingerers and non-malingerers.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to my Senior Supervisor, Dr. Jim Ogloff. Besides being one of the finest professors I have worked with, he has been tirelessly supportive throughout the sometimes difficult process of being a doctoral student.

I also wish to acknowledge the huge debt of gratitude that I owe to Dr. Allen Thornton. His assistance and support for this project was truly above and beyond the call of duty. I do not think that any thanks I can offer will ever be profuse enough.

I wish to thank my family, Mike, Frona, Bill, Glenn, Chimo, Boots, and Charlie, for their endless love and support. I am extremely lucky to have the best friends in the world, and I only hope that I can be as encouraging, helpful, and loving as they have been to me. Words cannot truly express the gratitude I feel toward these people.

Finally, I wish to thank (Dr.) Rob Saari for his support, advice, and love. He and his fun-loving spirit have given me a new perspective on life that has helped me more than I could have thought possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The Issue of Malingering in a Neuropsychological Context

The Diagnostic and Statistical Manual of Mental Disorders (4th edition; DSM-IV; American Psychiatric Association, 1994) defines malingering as "the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as . . . avoiding work [or] obtaining financial compensation . . .." It further states that

> malingering should be strongly suspected if any combination of the following is noted: (1) medicolegal context of presentation; (2) marked discrepancy between the person's claimed stress or disability and the objective findings; (3) lack of cooperation during the diagnostic evaluation and in complying with the prescribed treatment regimen, or (4) the presence of Antisocial Personality Disorder (p. 683).

It is important to differentiate malingering from other non-organic/medical disorders, including Factitious Disorder and Somatization Disorder or Conversion Disorder. Factitious Disorder involves the intentional production of symptoms, as in malingering, but the purpose differs: in Factitious Disorder, the patient wishes to take on the "sick" role for no apparent primary gain. Somatization Disorder and Conversion Disorder are diagnosed when no suggestion of intent or gain exists.

Even among those diagnosed as malingerers, some authorities perceive variability. Rogers (1997) conceptualizes malingering on a continuum, from those who produce completely falsified symptoms, to those who mildly exaggerate existing symptoms. Malingering and misrepresentation can take many forms (Faust & Ackley, 1998). A subject may put forth suboptimal effort, fabricate symptoms, exaggerate existing symptoms, mislead about causality, or may fail to report accurately on remaining

strengths. This variety makes identification of appropriate experimental and comparison groups for studies difficult.

Malingering as a strategy to obtain benefits is maintained by the fact that injury can be compensated through several means, including financial awards, evasion of civil or criminal liability, relief from work requirements, and the like. Nies and Sweet (1994) have noted that the fact that injuries can be financially compensated or otherwise rewarded virtually guarantees that malingering will occur at some rate. They assert further that this statement is supported by the finding that malingering is more rare in less developed areas (Miller & Cartlidge, 1972).

Malingering of head injury is a particularly difficult problem for neuropsychologists to deal with. It is almost presumable that neurological impairment will follow severe and possibly moderately severe traumatic brain injury, so the validity of impairment claims in such situations are often not an issue. Mild head injury is much more problematic. Comprehension of subtle neurological changes and their potential impact on neuropsychological processes resulting from mild head injury is limited and contentious. This, combined with the inability of objective imaging tools (magnetic resonance imaging, CT scans, EEG) to detect any such injury, diminishes clinicians' ability to state definitively when a patient is truthful in their claims of impairment (Levin, 1982). Clinicians can use information from records, interview, and other sources to assist them in drawing conclusions about the presence of malingering. However, this type of information is of varying reliability and validity. At the present time, neuropsychologists rely to a large extent on neuropsychological instruments and those tests designed specifically to identify malingering in addressing the validity of such injury claims. The

methods currently in use, and the problems in studying this issue, will be discussed in the following sections.

It is important to note first that symptom presentation following traumatic brain injury can include a wide range of neurological impairments, such as sensory and motor impairments, language difficulties, and, more rarely, visual deficits. By far, the most common impairment is in memory functions (Schacter & Crovitz, 1977), possibly because brain structures involved in memory are the most sensitive to oxygen deprivation, mechanical injury, and the like (Levin, Lilly, Papanicolaou, & Eisenberg, 1992). Also, laypersons seem to identify it most often as a typical symptom of brain injury or illness, and therefore it might be the most common symptom malingered overall (Williams, 1998). This review will focus for the most part on malingered memory impairment, as that is where the majority of research efforts have been directed.

General Issues in the Assessment of Malingering

Correct identification of questionable performance is essential. Accurate diagnosis of malingering not only limits unnecessary medical and insurance costs, it facilitates rapid and appropriate provision of medical and rehabilitative services. Considerable stigma is attached to a diagnosis of malingering, and such a diagnosis carries negative implications for the individual (Spreen & Strauss, 1998). For these reasons, false positive diagnoses should be avoided if at all possible.

One important factor contributing to accuracy of diagnosis is the base rate of malingering. Cullum, Heaton, and Grant (1991) describe the prevalence of neuropsychological malingering as a "significant minority" (p. 141). Reynolds (1998) stated that "reasonable and thorough research indicates that at least 25% of cases in

head injury litigation involve malingering" (p. viii). Other estimates of its frequency vary by setting. Malingering may occur in as many as 38% of disability claimants (Guilmette, Whelihan, Sparadeo, & Buongiorno, 1994). In a worker's compensation sample, Youngjohn (1991) estimated the rate of malingering to be 47%, and in personal injury cases, some research suggests it is as high as 64% (Heaton, Smith, Lehman, & Vogt, 1978). Pankratz and Binder (1997) concluded from their review of studies that 20 to 60% of mild head-injured patients with external incentives had unrealistically poor scores on neuropsychological instruments. The true rate of malingering is unknown, owing to the lack of a gold standard for assessing malingering, but it is evident that it occurs with some frequency, particularly in civil or criminal forensic situations. Assessment of malingering is therefore a very real concern, of which clinicians should always be mindful.

This being said, attention must be turned to the methods with which malingering can be identified. Prior to the 1980s, two assumptions were common: first, malingering was not prevalent, and second, it could be detected by astute clinicians. Several studies have evaluated the ability of psychologists to detect malingering of neuropsychological deficits on the basis of clinical judgement alone (Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, & Arkes, 1988; Goebel, 1983; Heaton et al., 1978; Trueblood & Binder, 1997). Methodological problems notwithstanding, clinicians were repeatedly shown to be unacceptably poor at identifying malingerers. Heaton and colleagues (1978), for example, found that the ability of 10 neuropsychologists to identify malingerers ranged between chance and 20% greater than chance. A more recent study (which therefore capitalizes on any effect of the recent growth of the literature on accuracy) reported that psychologists' classification "error rates" ranged from 0% to 25% across four cases of

identified clinical malingerers (Trueblood & Binder, 1997). Nevertheless, a sizeable body of literature has demonstrated the superiority of formal decision rules over subjective opinion in classification accuracy across a variety of situations (e.g., Dawes, Faust, & Meehl, 1989; Grove & Meehl, 1996). Thus, we must look beyond clinical judgement as the sole method of detecting malingering.

Six Strategies for Detecting Malingering

Rogers, Harrell, and Liff (1993) identified six potential strategies for detecting malingering on neuropsychological instruments. These strategies included (1) the floor effect, (2) symptom validity testing, (3) examination of performance curves, (4) examination of the magnitude of error, (5) inconsistency within or between evaluations, and (6) investigation of reported psychiatric symptoms.

The "floor effect" entails administering tests on which even severely impaired individuals would be expected to perform well. However, the possible malingerer is led to believe the opposite, thus cueing him or her to do worse than cognitively impaired subjects and identifying him or herself as malingering. The Rey 15-Item Test is an example of such a strategy (Lezak, 1994). It involves presentation of 15 simple, redundant items arranged in five rows of three items. The subject is told that it is a memory test of 15 items, with emphasis on the implication that it is a difficult test. The subject views the card for 10 seconds, and then reproduces as many items as he or she can recall. It is easy and quick to administer and score. Several researchers have reported on the success of this instrument in distinguishing between brain-injured patients and malingerers (Bernard & Fowler, 1990; Bernard, Houston, & Natoli, 1993; Guilmette, Hart, Giuliano, & Martin, 1993; Lee, Loring, & Martin, 1992; Millis & Kler,

1995). One shortcoming of the test is that the use of a cutoff score (e.g., number of items recalled correctly) may not be the most valid or reliable way of discriminating between malingerers and brain-injured patients. Wiggins and Brandt (1988) found that malingerers' performance on the Rey 15-Item was quantitatively similar to that of patients, but qualitatively there were notable differences. Other studies have found a significant relationship between performance on this test and IQ or age (Hays, Emmons, & Lawson, 1993; Schretlen, Brandt, Krafft, & VanGorp, 1991; Simon, 1994). In hopes of addressing some of these issues, Griffin, Glassmire, Henderson, and McCann (1997) redesigned the Rey, calling the new test the Rey II. Although they found positive results, this version of the test has not yet been thoroughly tested in the literature.

Pankratz (1979, 1983) was one of the early proponents of the second approach for detecting malingering, symptom validity testing (SVT). Generally, this method involves presentation of individual stimuli, such as a string of digits, followed by a recognition trial in which the subject is forced to try to choose the previously presented stimulus from a set of at least two options – hence the alternative name for this strategy, "forced choice testing." This procedure is based on recognition memory. Memory research has demonstrated that recognition memory is preserved even among persons who have bona fide memory deficits or who have suffered a head injury (Brandt, Rubinsky, & Lassen, 1985; Haines & Norris, 1995). Furthermore, binomial distribution theory suggests that even severely impaired individuals responding honestly (i.e., operating at chance levels) will be correct in their recognition at least 50% of the time ina two-choice task. Therefore, if the subject's accuracy rate falls below 50%, the only possible explanation is that he or she was attempting to appear more impaired than s/he actually is (Sinnett & Holen, 1995). However, evidence from clinical and research use of

this procedure indicates some flaws in this reasoning. First, performance above chance levels on forced-choice procedures does not rule out the possibility of malingering. Nies and Sweet (1994) suggested that very high-functioning malingerers may recognize that extremely poor performance (i.e., below chance) is improbable, and thus may score below expected levels but not below chance (Binder & Willis, 1991; Guilmette et al., 1993; Slick et al., 1994). Therefore, efforts have recently been directed toward identification of "below expected levels" on forced-choice procedures. For example, an entire sample of brain-injured patients with no external incentive performed above 54% correct on the Portland Digit Recognition Test (PDRT; Binder & Willis, 1991). This prompted Binder (1992) to recommend a cutoff of 54% as the most accurate cutting score for the assessment of malingering with this test. Second, scores below chance can be produced for reasons other than malingering, including poor motivation in the absence of potential secondary gain. In a related vein, Trueblood & Binder (1997) found that SVT data did not improve clinicians' accuracy in identifying clinical malingerers. Other tests using this method include the Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Spellacy, 1996), the Test of Memory Malingering (TOMM; Tombaugh, 1996), the 21-Item Test (Iverson, Franzen, & McCracken, 1991), and Warrington's Recognition Memory Test (RMT; Warrington, 1984). Note that all of these tests, with the exception of the RMT, were designed specifically to evaluate malingering; the RMT is typically classified as a purely neuropsychological instrument.

The third detection method is known as the performance curve strategy (Rogers et al., 1993). This approach takes into account the fact that honestly responding individuals are likely to perform in a manner reliably different from brain-injured persons. For example, malingerers might fail easier items on a test, while passing more difficult

items. Frederick and Foster (1991) modified the Test of Nonverbal Intelligence (TONI) into a forced-choice format. They found that measures of simulators' performance curve and consistency were significantly different from those of patients. Performance on early sets of Raven's Matrices compared to later (harder) sets was found to distinguish between malingerers and controls (Gudjonsson & Shackleton, 1986). This method has also been used with neuropsychological tools including the Rey Auditory Verbal Learning Test (RAVLT; Bernard, 1991) and the California Verbal Learning Test (CVLT; Millis, Putnam, Adams, & Ricker, 1995), the Wisconsin Card Sorting Test (WCST; Bernard, McGrath, & Houston, 1996), the Wechsler Memory Scale-Revised (WMS-R; Mittenberg, Azrin, Millsaps, & Heilbronner, 1993), the Digit Span subtest of the Wechsler batteries (Greiffenstein, Baker, & Gola, 1994), the Paced Auditory Serial Addition Test (PASAT; Strauss, Spellacy, Hunter, & Berry, 1994), and the Category Test (Tenhula & Sweet, 1996).

Examination of the magnitude of error (either qualitative or quantitative) in test results is the fourth malingering identification strategy. It is assumed that honest participants who do not know the correct answer to a question will not respond absurdly, nor will they give "near miss" responses. Malingerers are not expected to know this, and therefore may answer in these manners. Rawling and Brooks (1990) developed a Simulation Index based on simulators' qualitative errors on the WAIS-R and WMS-R. Although Rawling (1992) found positive results upon cross-validation, Milanovich, Axelrod, and Millis (1996) found that it performed poorly, especially with brain-impaired participants of various diagnoses.

Lezak (1995) and Cullum, Heaton, and Grant (1991) have asserted that inconsistency, either among an individual's test scores or between test results and real-

world performance, may be the hallmark of malingering. This is the basis of the fifth detection approach, the evaluation of atypical presentation. Recent research has examined the consistency of results across multiple testings (Beetar & Williams, 1994; Reitan & Wolfson, 1996). Reitan and Wolfson (1996) developed a Dissimulation Index, based on difference scores calculated from a subject's performances on two administrations of WAIS-R and Halstead-Reitan tests. They observed that litigants' results were much less consistent than nonlitigants'. On the negative side, Pankratz and Binder (1997) noted that marked inconsistency in performance is sometimes characteristic of persons with brain injury. Greiffenstein and colleagues (1994) have set out specific criteria for the detection of malingering using observations from multiple sources of data. The criteria include: (1) two or more scores in the "severe" range relative to normative groups on neuropsychological tests; (2) an improbable symptom presentation contradicted by records or observation; (3) claims of total disability in a major social role more than one year postinjury; and (4) claims of remote memory loss. Greiffenstein's group (1994) suggests that a diagnosis of malingering is probable if the subject meets two or more of these criteria. Tombaugh (1997) concluded that this approach has limited clinical utility because of the overlap between valid and invalid patterns of performance, as well as the rarity of classic signs of invalidity.

The final strategy for the detection of malingering capitalizes on the assumption (and support from some studies) that laypersons have limited or inaccurate knowledge of actual psychological sequelae of brain injury (Aubrey, Dobbs, & Rule, 1989). Thus, malingerers are expected to endorse improbable or bizarre symptoms. Consistent with this theory, Heaton and colleagues (1978) observed that malingerers produced elevations on MMPI scales that were significantly higher than profiles produced by truly

brain-injured participants. As a basis for differentiating truly brain-injured patients from malingerers, some researchers have attempted to capitalize upon the complexity of memory and the impact of brain insult upon memory. Malingerers are expected not to appreciate that complexity (Baker, Hanley, Jackson, Kimmance, & Slade, 1993; Horton, Smith, Barghout, & Connolly, 1992; Wiggins & Brandt, 1988). In contrast, some studies have found that laypeople do have an accurate understanding of the sequelae of brain damage (Mittenberg, DiGiulio, Perrin, & Bass, 1989). Similarly, some researchers have concluded on the basis of their results that one should not expect to identify malingering by exaggerated poor or bizarre performance (Bernard, 1990; Wong, Regennitter, & Barrios, 1994). Other research has evaluated the impact of information about symptom presentation on ability to elude detection as a malingerer (Frederick & Foster, 1991; Hayward, Hall, Hunt, & Zubrick, 1987). MMPI-2 research has discovered that knowledge about strategies for escaping detection is more effective in achieving that goal than is information about typical symptomatology (Faust & Ackley, 1998; Lamb et al., 1994; Rogers, Bagby, & Chakraborty, 1993; Wetter, Baer, Berry, & Reynolds, 1994).

Summary. To date using traditional statistical design and analysis, none of the methods described above has been found to be reliably more accurate than the others in detecting malingered head injury. Research efforts directed at finding the best method have been hampered by methodological weaknesses, or differences that preclude direct comparisons of results. Meta-analysis allows combination and comparison of results from many different types of studies. Some approaches (such as symptom validity testing) appear more promising than do others, but until results of studies are standardized and study design characteristics are taken into account, no firm conclusions may be drawn. The following section will examine the methodological

problems faced by malingering researchers. This will assist in determining how study characteristics will have to be dealt with in a meta-analysis.

Methodological Issues in Research on Malingering

Malingering research, as with other areas of study, is faced with methodological pitfalls. These have been addressed to differing degrees of success by some researchers. Unfortunately, no study can completely deal with every problem, so results of separate studies have to be combined in some way to obtain a more complete picture of malingering. Meta-analysis is the ideal tool for achieving this goal (Glass, McGaw, & Smith, 1981), and for this reason, the present investigation was undertaken. In order to highlight issues for this meta-analysis to examine, we must identify the problems with the extant literature.

Studies in this area generally fall into one of two categories: simulation or analogue studies (labeled here as "experimental" studies), and those using the known-groups design ("clinical" studies). The primary issue with regard to the experimental/ simulation paradigm is its lack of generalizability to real clinical situations. Tests with very good accuracy in detecting malingerers in simulation experiments have failed when applied to clinical malingerers (Rogers et al., 1993). Analogue studies present the paradoxical predicament of "[asking] subjects to comply with directions to fake in order to study those who fake when asked to comply" (Rogers et al., 1993, p. 257). These studies typically involve normal individuals who are asked to imagine how a brain-damaged person would perform. This is potentially a difficult role to imagine for a person who may never have encountered a brain-injured individual, or who is completely naïve regarding the typical sequelae of head injury. Pankratz and Binder (1997) noted

that simulation research is based on the assumption that malingerers are normal people who cheat, which he criticized as a flawed assumption. Moreover, research demonstrates that 10 to 20% of participants are unwilling or unable to malinger, even when given incentives (Pankratz & Binder, 1997). And in spite of such incentives, participation in a research project is different from situations that elicit clinical malingering (Beetar & Williams, 1994). It is believed that at least some clinical malingerers are individuals who have been injured, but are exaggerating the magnitude of their impairment.

Alternatively, some malingerers may not have been injured, but have been thoroughly coached in presenting as brain-damaged and escaping detection. It is also possible that some malingerers may be very skilled at deceiving; this possibility is the basis for the inclusion of Antisocial Personality Disorder (APD) in DSM-IV as a condition where malingering should be considered. Pankratz and Binder (1997) have suggested that If this is true, then university students would not be a very valid sample of participants for malingering studies, as the base rate of APD in students is low. However, as it is unlikely that all malingerers meet criteria for APD, the use of non-APD participants is justifiable.

Finally, because it is almost impossible to verify when malingering has occurred, the similarity between performances of true clinical malingerers and simulators is difficult to establish. For example, some researchers assume that clinical malingerers would fail all tests of suboptimal effort, and this is likely how many simulated malingerers perform in studies. At least one case reported in the literature, however, challenges this notion. The patient described by Palmer, Boone, Allman, and Castro (1995) was a verified malingerer who failed both Dot Counting (Lezak, 1983) and the Rey 15-Item, but not the

Word Recognition Test (Lezak, 1983). Several studies have attempted to address the issue of coaching (e.g., Coleman, Rapport, Millis, Ricker, & Farchione, 1998; Hiscock, Branham, & Hiscock, 1994; Lamb, Berry, Wetter, & Baer, 1994; Rose, Hall, & Szalda-Petree, 1995), but the thoroughness of the preparation the participants receive is likely limited in comparison to the coaching a litigating individual might be expected to receive from a dedicated lawyer. A more realistic comparison group would involve asking non-compensation-seeking brain-damaged participants to exaggerate existing or rehabilitated difficulties.

Validity problems also exist with regard to use of "known groups" – clinical malingerers who have been caught or otherwise identified. Some studies have used such a sample as the experimental group, in hopes of increasing generalizability (e.g., Greiffenstein, Gola, & Baker, 1995; Suhr, Tranel, Wefel, & Barrash, 1997). The first major problem is how such a group would be identified, given the lack of validity and reliability in assessing malingering. People involved in litigation or who are seeking compensation for injuries have been identified as being at increased risk for exaggerating or malingering. The major shortcoming of this method is the fact that simply pursuing compensation is insufficient grounds for labelling a participant as a malingerer; not all of those who seek compensation are malingerers. This approach is also unhelpful to clinicians working in workers' compensation or insurance settings, who are asked to identify malingerers from a sample of litigants. Moreover, it can be argued that malingerers who have been caught might differ systematically from malingerers who are not caught (Faust & Ackley, 1998).

Other studies have classified participants as probable malingerers on the basis of questionable results on neuropsychological or malingering tests. The reasoning behind

this tactic is somewhat circular, as it assumes that tests can identify malingering in the first place. This leads to the problem that solid interpretations of differences between groups cannot be made even when they occur in the expected direction. For these reasons, researchers have been cautioned against using these strategies alone in malingering research (Rogers et al., 1993).

Methodological difficulties persist no matter how close or far studies are from real-world scenarios. This precludes making firm conclusive statements about the effectiveness of malingering detection strategies on the basis of traditional literature review methods. Fortunately, such a situation is exactly why meta-analysis was developed (Glass et al., 1981). Rogers and colleagues (Rogers, Gillis, Dickens, & Bagby, 1991) suggest that this issue is best dealt with by using a combination of the simulation and known-groups approaches; that is, potential malingering indices should initially be tested under well-controlled situations using the simulation paradigm. Instruments that show promise can then be evaluated in clinical settings.

Another criticism of simulation studies is the questionable value of the comparison groups being used. Clinicians do not typically need to distinguish between normal subjects performing at an optimal level and those who report injury but are actually malingering (Faust & Ackley, 1998). However, this comparison is often used in studies of malingering. Choice of an appropriate comparison group is essential in useful assessment of malingering detection. Specifically, identification of malingering requires comparing groups who have similar magnitudes of injury but have differing motivations for performance (Tombaugh, 1997).

Finally, unlike specific malingering tests (e.g., the Rey 15-Item Test, the Portland Digit Recognition Test), neuropsychological instruments were not designed to assess

malingering. However, they are sometimes used in studies to try to distinguish between malingerers and non-malingerers. The validity of these instruments for this purpose may be limited, but that has not yet been established.

A recent review of the literature on malingering in neuropsychological assessment concluded that 50 years of research has not led to a consensus on appropriate assessment of malingering (Nies & Sweet, 1994). It is important to note, however, that most of the research, particularly well-designed and -conducted research, has only been done in the past 15 years or so (Pankratz & Binder, 1997). Some of the issues raised in the present review have been addressed in the literature, but generally in a piece-meal fashion. A thorough, quantitative review of the effectiveness of efforts to detect malingering on neuropsychological instruments is necessary to sort through the confusing outcomes. It is apparent from the present survey that the research corpus has grown to an extent that a meta-analysis would serve the area well. This meta-analysis will attempt to address the issues raised in the preceding review.

Research Questions and Hypotheses

This review of the literature has highlighted many issues in the area of neuropsychological assessment of malingering, but three questions are especially prominent:

1. Do individual clinical neuropsychology instruments differ in the degree to which they can detect malingering?

   It was expected that some individual instruments would be more effective than would others at identifying malingering. That is, the distribution of effect sizes obtained by these instruments with malingerers would be significantly different from

zero (i.e., the mean score obtained with normal or brain-injured samples). In other words, these instruments would discriminate better between malingerers and non-malingerers (brain-injured or controls) than would other instruments. The reason for this is that although clinicians know that these instruments are highly sensitive to brain damage, malingerers may not realize this, and may not adjust their strategies accordingly (i.e., they may exaggerate their poor performance). Ideally, however, even if malingerers did adjust their performance strategies, a truly sensitive test would detect their exaggeration.

It was expected that mean scores obtained by malingerers would typically be in the more pathological direction, whereas brain-injured patients or normal controls would in the less pathological direction. For example, malingerers usually do not realize that even severely impaired persons can do well on the Rey 15-Item test, and tend to obtain very low scores on this test. Brain-damaged individuals and controls, on the other hand, score within normal limits.

2. Grouping clinical neuropsychology instruments together by function, are some domains of cognitive functioning (i.e., memory; visuospatial (e.g., Rey Complex Figure); language (e.g., Word Fluency); executive functions (e.g., WCST)), or malingering detection strategies (i.e., symptom validity testing) better able to detect malingering?

It was hypothesized that some domains of cognitive functioning or malingering detection strategies would more clearly distinguish between malingerers and non-malingerers than others. That is, compared to less effective cognitive domains or malingering detection strategies, the effect sizes associated with the more successful cognitive domains or strategies would be larger and the confidence intervals for

these domains would not overlap with zero. Specifically, it was expected that effect sizes using malingering tests would typically be larger than those of other cognitive domains.

3. Is there an interaction between categories of tests and categories of participants?

It was hypothesized that in studies using a brain-injured comparison group, naïve and coached malingering groups, and tests from categories believed to be most sensitive to malingering, such an interaction would exist. Thus, using symptom validity testing, naïve malingerers were expected to show the most pathological scores, followed by coached malingerers. Litigants and clinical participants suspected of malingering would show the least pathological scores on these tests, but it was expected that the effect sizes for these two groups would not be significantly different from each other. It was anticipated that all malingering group means would be significantly different from zero (i.e., the comparison group mean). Such a pattern would not be displayed by detection strategies or test categories that were not as sensitive to malingering, such as language or executive functions. That is, it was expected that there would be no effect size differences among malingering groups on these less sensitive cognitive domains or instruments.

METHOD

Studies

Studies were identified through several means. First, a search of computerized databases was conducted. PsycINFO (1967-1999), MedLine (1966-1999), Criminal Justice Abstracts (1968-1998), and Current Contents (1995-04/1999) were searched using the following terms: maling*, feign*, fake*, faking*, dissimu*, head injury, brain injury, neurop* or neuropsy*, and/or cogn*. Reference lists from the articles obtained through the computerized search were examined for other entries. Finally, a journal-by-journal search of *Archives of Clinical Neuropsychology*, *Journal of Clinical and Experimental Neuropsychology*, and *The Clinical Neuropsychologist* was conducted. These journals are those in which articles from the targeted domain are most frequently published. Dissertations were not included because of the limited information obtainable and the prohibitive cost associated with ordering these manuscripts. Posters and papers presented at conferences were not included because of the great difficulty associated with trying to obtain them, and the questionable representativeness of those presentations that could be obtained. Only studies published in English were included.

Studies selected for inclusion compared the performance of at least one group of malingerers (simulation or known-groups) with a control group of non-compensation-seeking brain-damaged patients on clinical neuropsychology instruments and/or tests designed to identify malingering. To be included, studies had to present sufficient information to calculate effect sizes (described below). Studies that focused on experimental measures, such as implicit memory (priming) tasks, were not recruited.

Similarly, studies using the MMPI or MMPI-2, which are not neuropsychological instruments, were not recruited. Finally, some instruments were investigated in only one study (e.g., the Kaufman Hand Movements Test). When possible, these instruments were categorized into cognitive domains and their effect sizes were evaluated. If it was not possible to categorize the instrument into a particular domain, the study was excluded.

Studies selected for inclusion compared the performance of at least one group of malingerers (simulation or known-groups) with a control group of either non-compensation-seeking brain-damaged patients or normal, healthy controls on clinical neuropsychology instruments and/or tests designed to identify malingering. To be included, studies had to present sufficient information to calculate effect sizes (described below).

Procedure

One rater (the author) coded all studies selected for inclusion. One potential concern about this was that the rater's coding of variables would shift with time and experience. To control for this coding "drift," the first 10, middle 10, and last 10 studies coded were re-coded upon completion of coding. Any discrepancies found by this procedure were corrected by recoding all relevant studies. A coding sheet (see Appendix A) was used for all studies.

Demographic information. Several variables were coded to assess the relationship between participant demographic variables and ability to detect malingering. These included (where appropriate): age; education; percentage of the sample that was

male; diagnosis; criteria for injury; time since injury; and sample characteristics (i.e., university students, community-dwelling adults, etc.).

Methodological information. Study design variables were coded to evaluate the effect of study design quality on malingering detection ability (effect sizes). These included (where appropriate): random selection; random assignment; consecutive referrals; matching on demographic factors; participant exclusion information; malingering incentive; participant instructions (including coaching); and presence of a manipulation check. Study quality was characterized as "methodologically sound" on the basis of presence of any of the following: random selection or assignment, consecutive referrals, matching on demographic variables, or manipulation check. Quality was coded as "other" on the basis of absence of all of the same characteristics. Quality was dichotomized to deal with the problem that some of these characteristics were not applicable for all studies. Also, the characteristics could not be weighted equally. Therefore, these characteristics could not be combined algebraically, so straightforward dichotomization was used.

Cognitive Domains. Individual neuropsychological tests were clustered into cognitive domains to examine whether general categories of cognitive functioning or types of tests were more or less able to discriminate between malingerers and non-malingerers. Factor analyses exist in the literature evaluating the construct validity of common neuropsychology instruments, especially the Wechsler Memory Scale-Revised and other memory tests. Following the work by Leonberger, Larrabee, and their colleagues (Leonberger, Nicks, Larrabee, & Goldfader, 1992; Larrabee & Curtiss, 1995; Larrabee, Kane, Schuck., & Francis, 1985), the instruments used in the present study were collapsed into cognitive domains, as presented in Appendix B. Where tests used

in the present study were not included in the published factor analyses, they were categorized into cognitive domains following Lezak (1995). Tests were categorized as tests of malingering if they were designed specifically for that purpose, according to articles in which they first appeared or manuals accompanying the instrument. It is recognized that "malingering" is not a cognitive domain, but it is included here in the general category of "cognitive domains" to simplify description. It is also recognized that although the clustering of instruments into domains is based here upon factor analytic studies, any grouping of neuropsychological tests is somewhat arbitrary because few if any of these tests tap only one domain of cognition.

Clinical neuropsychology instruments and scores. A list of instruments and associated scores used in this meta-analysis is shown in Appendix B.

Subjects

Comparison Groups

Brain-Injured Individuals. One of the groups to which malingerer groups were compared was subjects with documented traumatic brain damage. Such brain injuries often resulted from motor vehicle accidents, gunshot wounds, or falls. Many studies used comparison groups that included a wide range of head injury severities, from no loss of consciousness (LOC) to months of coma. To control for potential statistical noise or error introduced by comparison group variability, studies were selected such that all comparison groups had a head injury of moderate severity [i.e., Glasgow Coma Scale score of 9 to 12 upon admission to hospital, or loss of consciousness from 20 minutes to 6 hours (Lezak, 1995, p. 755)]. In some studies, subjects were described in terms of the numbers of participants who had certain durations of LOC. In these cases, the author

tallied the numbers falling into mild, moderate, and severe categories of LOC. The category associated with the median number of participants was chosen to represent the study. Participants were drawn equally from in- and outpatient samples. Studies that did not include a comparison group with documented head injury were excluded from the analyses. Also excluded from analyses were studies in which comparison group participants were seeking compensation or in litigation at the time of participation in the study. A list of all studies excluded from analyses, and the reason for exclusion, is provided in Appendix C.

Normal Healthy Controls. One difficulty with comparing malingerers to brain-injured individuals was the heterogeneity associated with such a group. This was partially addressed by limiting brain-injured samples to those involving moderate head injury. However, these participants still had injuries in a wide variety of brain areas, thus potentially introducing an element of method error into the analyses.

Moreover, norms tables used by neuropsychologists to evaluate cognitive performance compare a given group to normal individuals. It was thought that such a comparison would be clinically useful here. A normal control group would provide a consistent standard against which malingerers could be compared. Malingerers were therefore also compared to a less heterogeneous sample, a group of healthy normal controls.

Malingerer Groups

Naïve Malingerers. The group labelled Naïve Malingerers was comprised of university students or community-dwelling adults. Several studies screened these participants for history of head injury (32% of studies), neurological disorder (30%),

substance abuse (12%), or psychological problems (21%). These participants were given no information about how (or sometimes, what disorder) to malinger, or what strategies to use to avoid detection.

Coached Malingerers. The Coached Malingerer group was also comprised of university students or community-dwelling adults. The majority of these participants had been screened for history of head injury (90%), neurological disorder (87%), substance abuse (44%), or psychological problems (50%). Coached malingerers were given instructions either on how to malinger head injury or memory impairment, or on how to avoid detection (e.g., pass easy items, but fail harder ones; do not respond inconsistently; etc.).

Litigant Group. The Litigant group was comprised of persons who had allegedly suffered a head injury, and were either seeking compensation (e.g., from a worker's compensation organization) or in litigation for damages. In many cases, no documented evidence of head injury (such as duration of loss of consciousness, Glasgow Coma Scale scores, positive results of neuroimaging) was available. In the majority of instances (54% of studies), when a description of the type of injury was available, participants had suffered at most a mild head injury. Of the studies in this group, 11% excluded participants if they had a history of neurological problems (other than any resulting from the head injury), 21% excluded participants for psychological problems, and only 2% of studies excluded participants if they had a history of substance abuse.

Suspected Malingerer Group. The Suspected Malingerer group was similar to the Litigant group. Of the studies in this group, 77% were comprised of participants who had suffered a mild head injury. In addition, however, these participants had performed in such a way outside of the testing situation as to lead examiners to suspect that they

were malingering. Studies in this category considered participants to be possibly malingering if (a) their subjective complaints of distress or disability far exceeded what would be expected given their medical history, and (b) their pattern or level of neuropsychological performance was highly unusual (e.g., characterized by marked inconsistency in performance, or unrealistically poor performance), if (c) they were pursuing compensation or in litigation. Of these studies, 18% excluded participants if they had a history of neurological problems (other than from the head injury). None of the studies in this group stated whether participants had been excluded for psychological problems or substance abuse.

Effect Size Analyses

Calculation of effect sizes depends on the quality of the information used in the analyses. Therefore, it is essential to carefully screen studies to determine whether they meet criteria for inclusion in the study. In this investigation, a total of 102 studies were carefully examined and evaluated for inclusion. Forty-nine studies were excluded because (a) effect sizes could not be calculated using the data provided (e.g., did not present standard deviations associated with mean scores; no control group; missing data), or (b) using instruments that were not in clinical use (i.e., experimental measures such as priming memory tasks) or that failed to fit into any particular cognitive domain (e.g., MMPI). This left 53 studies, comprising 66 separate samples, for further investigation. Appendix C lists the studies that were included and excluded and the reasons for excluding the studies.

Effect sizes were calculated using the remaining 66 samples. Effects were pooled for individual studies that used the same participants or instruments more than

once. Occasionally, more than one study used the same sample (either in part or in total). To correct for the redundancy of these samples, and to avoid one sample exerting an undue influence on the overall effects, the study reporting on the largest number of participants was used.

Effect size calculation. An effect size estimate was calculated for each test score comparison using DSTAT (Johnson, 1989). This produced an effect size estimate, Hedges' *g* (Hedges & Olkin, 1985), which is defined as the difference between the mean scores of the two groups (malingering group mean minus control group mean), divided by the pooled standard deviation. In the cases that studies did not present means and standard deviations, effect sizes were estimated from available data, such as *F* ratios (two groups) or *p* values.

The estimate *g* is influenced by the participant sample size; it overestimates the true population effect size when used with smaller sample sizes. Therefore, an unbiased effect size estimate, Cohen's *d*, was calculated for each test score comparison, also using DSTAT. The variance of each *d* value was calculated. The variance was then used to compute *w*, a weighting factor for the unbiased effect size *d*. The weighted *d* was symbolized as *wd*. The sum of the individual *wd*'s was divided by the sum of the individual *w*'s to produce *d.*, the weighted average composite unbiased effect size. The variance of this statistic was used to establish 95% confidence intervals around *d*. (Thornton & Raz, 1997; Hedges & Olkin, 1985).

Small cell sizes in a meta-analysis produce results that are unreliable. However, there is no consensus among meta-analytic researchers as to the minimum number of samples or studies necessary for reliable effect size values. Taking into account the number of samples available for analysis in this study, therefore, a minimum number of

four samples per cell was selected for this study. Statistical comparisons were not conducted for any categories with fewer than four samples.

Examination for outliers. After calculating effect sizes, the next step was to examine the overall distribution of effect sizes to identify and remove any potential outliers in the data. Removal of outliers is necessary since outliers can artificially inflate or deflate the mean effect size, which is the best estimate of the true population effect under investigation. The removal of outliers in this study was achieved by examining histograms of the effect sizes of all the included studies. Effect sizes that were clearly located beyond the tails of the distributions were identified and excluded from analyses. These effect sizes were in all cases associated with a Z score greater than 3.00.

Homogeneity of the effect size. Homogeneity of the effect size, H_T, was calculated using DSTAT to test whether the studies had a common underlying effect, or whether the studies possessed different characteristics that influence effect size. If the homogeneity of variance test H_T was not rejected, this would indicate that the variability among effect size estimates was consistent with a single underlying effect, and that the weighted mean effect was a good estimate of the population effect size, theta. If the homogeneity of variance test was rejected, this would suggest that subgroupings of studies exist that possess different characteristics, causing them to differ in effect size. In other words, the studies (or groupings of studies) are like apples and oranges. In such a case, this was then followed with the between-group homogeneity statistic, H_B, to determine which groups differed from each other. Significance tests determined the probability of Type I error associated with H_T, H_B, and H_W, which are distributed as χ^2 with (k – 1) degrees of freedom.

The goal of the next part of the procedure was to determine methods of partitioning the groups to minimize the heterogeneity, if possible. Correlations and chi-square analyses were conducted to assess the relationship between potential moderator variables (e.g., study quality; malingering incentive; participant characteristics) and effect size. These procedures provided some suggestions for breaking the larger groups into smaller subgroups. The smaller groups were then run through the entire procedure again to determine whether any decrease in heterogeneity was achieved.

Publication bias. Another important consideration in performing a meta-analytic investigation is determining whether a significant number of studies are not available for review because of small effect sizes and/or small sample sizes. Because there is a bias against publishing studies with small effects or small sample sizes, it is likely that there are unpublished studies not available for inclusion in the meta-analysis (Rosenthal, 1991). This poses a challenge for the meta-analyst since it makes it nearly impossible to obtain a truly representative sample of all studies conducted in an area. This tends to artificially inflate the effect sizes observed in meta-analyses.

Meta-analytic researchers have developed techniques to identify and correct for publication bias. One method is to calculate the number of studies with null effects that would be required to negate the overall effect size (Rosenthal, 1991). Another method is to examine funnel plots (Light & Pillemer, 1984). The purpose of a funnel plot is to detect whether studies have not been published as a result of having small or no effects with small samples. These methods were used in the present investigation.

RESULTS

The central research question of this meta-analysis was whether different domains of cognitive functioning demonstrated differential ability to distinguish between malingerers and non-malingerers. To address this issue fully, several additional analyses were necessary. The goal of these additional analyses was to examine the distribution of effect sizes for outliers, to determine whether publication bias was operating, to inspect the degree of heterogeneity associated with the observed effect sizes, and to identify any variables that moderated the observed effect sizes. The question of whether different cognitive domains differentially discriminated between malingerers and non-malingerers will be addressed in the final section of the Results.

Initial Inspection of Effect Size Distributions

Figure 1 presents the distribution of test score differences (i.e., effect sizes) between malingerers and brain-damaged controls. Notice that some effect sizes are very large and lie beyond the tail of the normal distribution. These effects (see Appendix C) were judged to be outliers and were removed from analyses. All outliers were associated with a z-score greater than 3.00. The mean, median, and modal effect sizes appear to fall around $d = 1.00$.

Figure 1

Effect size distribution for malingerers vs. brain-injured participants (aggregated by sample).

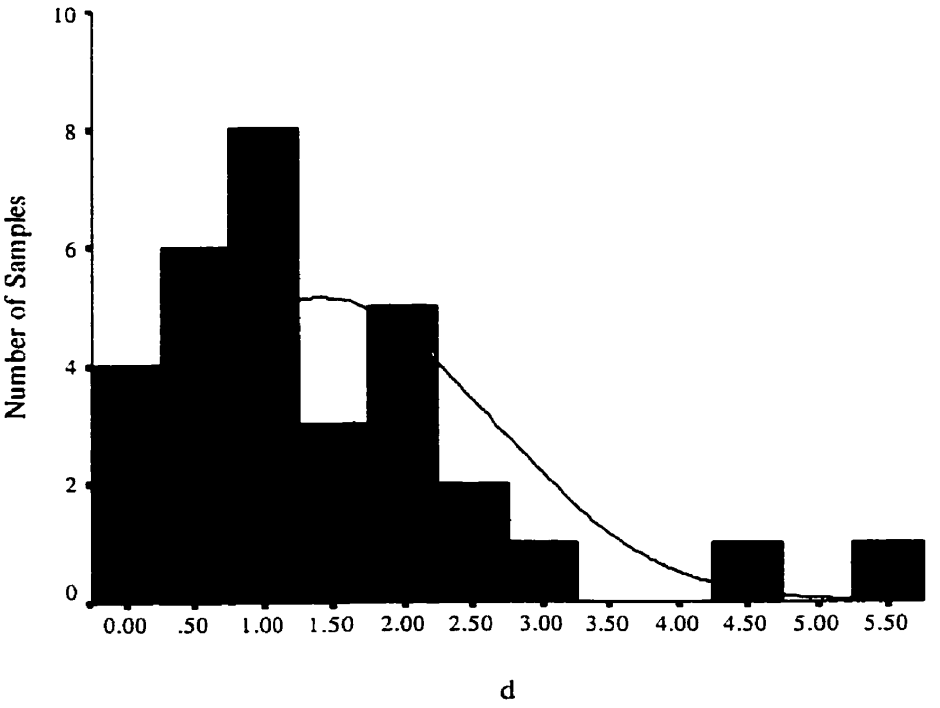
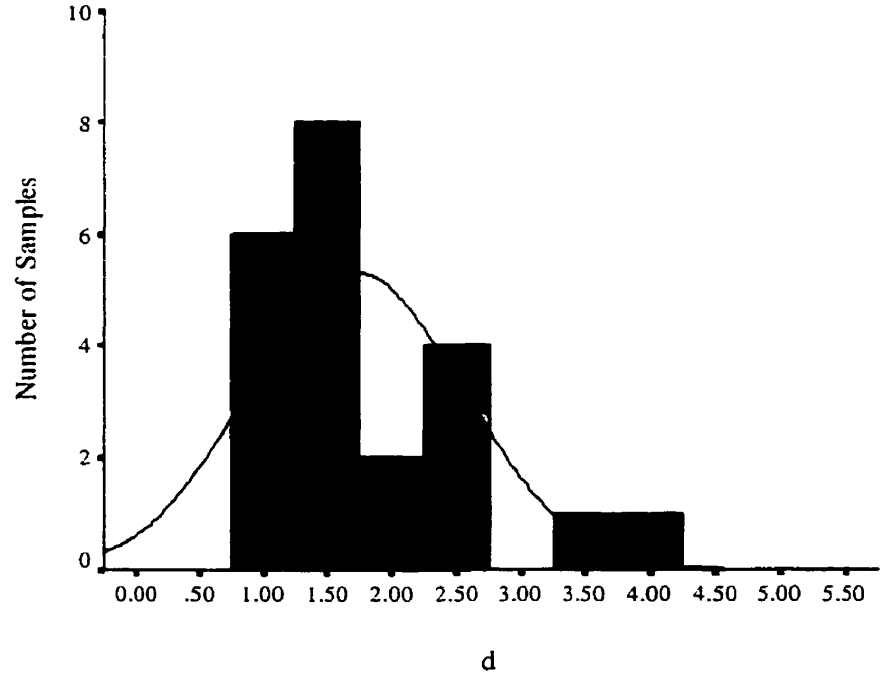


Figure 2 presents the distribution of effect sizes for malingerers compared to healthy controls. Similar to Figure 1, some effects are large and separated from the rest of the distribution. These were judged to be outliers and were removed from analyses (see Appendix C). The median and modal effect sizes appear to fall around $d = 1.50$, which is larger than that observed in Figure 1. The exact values and statistical difference between them will be analyzed in subsequent sections.

Figure 2

Effect size distribution for malingerers vs. normal controls (aggregated by sample).



Publication Bias/The File Drawer Problem

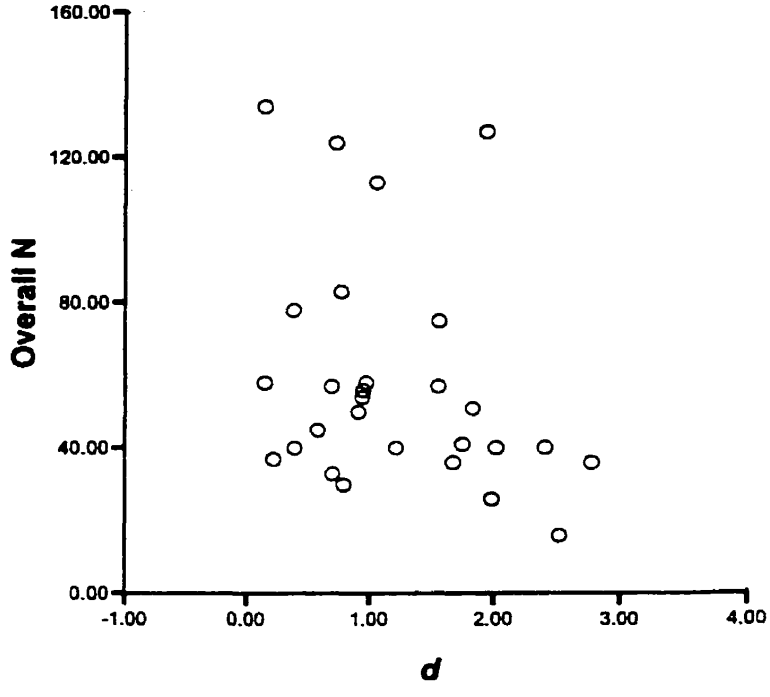
Publication bias exists when studies with small samples and small effects are not published (Rosenthal, 1991). Funnel plots can be used to detect whether studies have not been published as a result of having small or no effects with small samples. A funnel plot graphs the number of participants in a study against the overall effect size of the study. If publication bias exists, then on a funnel plot we would expect to see a gap in the plot where it approaches zero on the effect size axis (Light & Pillemer, 1984).

Brain-Injured Comparison Group. Figure 3 presents such a funnel plot for the studies using a brain-injured comparison group. It appears that there may be a gap near

zero, although it is small. This suggests the possibility that in this domain of neuropsychological or malingering research, some studies with null effects and small sample sizes have not been published.

Figure 3

Funnel plot of sample size by effect size (d) aggregated by sample: Brain-injured comparison group.



Although the funnel plot suggests that publication bias may be operating in this domain of research, given the magnitude of the observed effect and the power associated with these analyses (presented in subsequent sections), it is likely that many studies with null effects would be necessary before the observed effect would become not significantly different from zero. Therefore, it is unlikely that these theoretically

unpublished studies would have had a significant impact on the malingering effect observed in this meta-analysis.

In order to determine the number of studies with null effects that would be necessary to bring the observed effect using the brain-injured comparison group to approximately zero, the "fail-safe *N*" (Rosenthal, 1991) was calculated. However, given that the data appeared to be positively skewed (see Figure 1), the mean effect size would not accurately represent the central tendency of the data. Therefore, the median effect size was substituted for the mean in the calculation. To bring the observed median effect size from .95 (see Table 1) to .05, 342 studies with null effects would be necessary. It is unlikely that so many studies with null effects exist, and therefore, it is unlikely that the observed effect size is an artifact of selective sampling of studies (Hedges & Olkin, 1985).

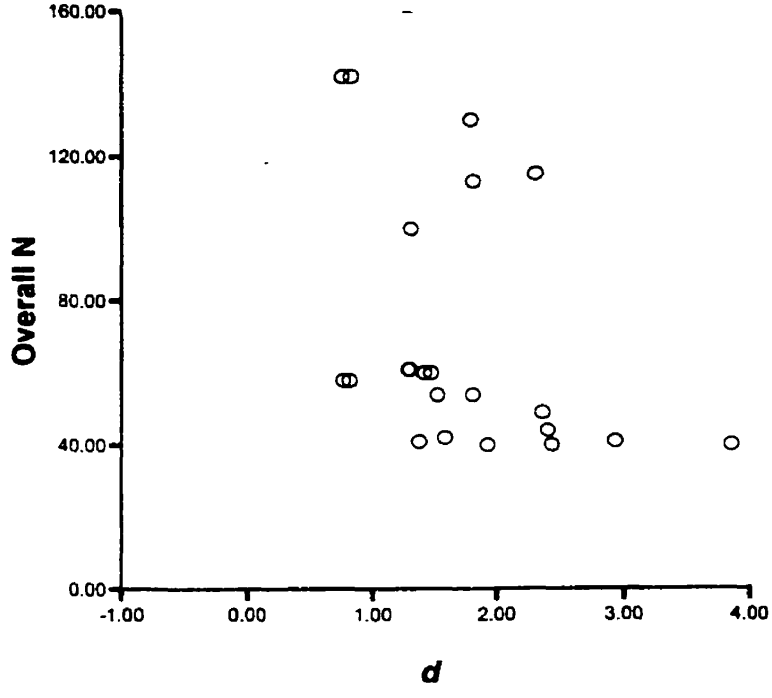
To further investigate the influence of sample size on effect size, overall sample size (i.e., malingering group *n* plus brain-injured group *n*) was correlated (Pearson's *r*) with effect size. It was anticipated that larger effect sizes would be associated with smaller sample sizes, and therefore a one-tailed test was used. Consistent with expectations, the resulting correlation was significant ($r = -.38$, $p = .02$).

This indicated that sample size exerted an undue influence on effect sizes, and therefore the group of effects was divided into two groups, large *n* studies and small *n* studies. Given the skewness of the distributions, the median overall *n* of 51 was chosen as the splitting point. For studies with sample sizes greater than or equal to 51 ($n = 14$), $d = .76$ (95% CI = .63 – .88), whereas for samples smaller than 51 ($n = 15$), $d = 1.28$ (95% CI = 1.09 – 1.48). These effect sizes were significantly different ($\chi^2 = 20.71$, $p < .001$).

Normal Control Comparison Group. Figure 4 presents a funnel plot for the studies using a normal control comparison group. It appears that there may be a gap near zero. This suggests the possibility that in this domain of neuropsychological or malingering research, some studies with null effects and small sample sizes have not been published.

Figure 4

Funnel plot of sample size by effect size (d) aggregated by sample: Normal control comparison group.



As with the brain-injured comparison group, the funnel plot suggests that publication bias may be operating in this domain of research. However, given the magnitude of the observed effect and the power associated with these analyses

(presented in subsequent sections), it is likely that many studies with null effects would be necessary before the observed effect would become not significantly different from zero. Therefore, it is unlikely that these theoretically unpublished studies would have had a significant impact on the malingering effect observed in this meta-analysis.

The "fail-safe *N*" (Rosenthal, 1991) was calculated for the normal control group, again using the median effect size in place of the mean. To bring the observed median effect size from 1.55 (see Table 1) to .05, 450 studies with null effects would be necessary. Again, it is unlikely that so many studies with null effects exist, and therefore, it is unlikely that the observed effect size is an artifact of selective sampling of studies (Hedges & Olkin, 1985).

To further investigate the influence of sample size on effect size, overall sample size (i.e., malingering group *n* plus control group *n*) was correlated (Pearson's *r*) with effect size. It was anticipated that larger effect sizes would be associated with smaller sample sizes, and therefore a one-tailed test was used. Consistent with expectations, and similar to the findings with the brain-injured group, the resulting correlation was significant ($r = -.39$, $p = .04$).

This indicated that sample size exerted an undue influence on effect sizes, and therefore the group of effects was divided into two groups, large *n* studies and small *n* studies. Given the skewness of the distributions, the median overall *n* of 58 was chosen as the splitting point. For studies with sample sizes greater than or equal to 58 ($n = 12$), $d = 1.34$ (95% CI = 1.20 – 1.48), whereas for samples smaller than 58 ($n = 10$), $d = 1.69$ (95% CI = 1.49 – 1.88). As found with the brain-injured comparison group, these effect sizes were significantly different ($\chi^2 = 7.77$, $p < .05$).

Overall Effect of Malingering

Table 1 presents various indices of the central tendency of the effect sizes from the remaining studies. These have been broken down by comparison group, with malingerers compared first to brain-damaged individuals, and then to normal healthy controls. Indices of central tendency presented here are those recommended by Rosenthal (1991; 1995), including the unweighted overall mean and median effect sizes, and the proportion of studies producing positive effects (effect sizes in the expected direction). Rosenthal (1995) also recommended including the number of samples, the number of participants, and the median number of participants per study upon which the effect sizes are based; these values are also presented in Table 1. Stem-and-leaf plots for the two comparison groups are presented in Appendix C.

Table 1

Indices of Central Tendency for Remaining Effect Sizes

Index	Comparison Group	
	Brain-Injured vs. Malingerers	Normal Controls vs. Malingerers
Unweighted Mean Effect Size (*SD*)	1.17 (.75)	1.72 (.76)
Unweighted Median Effect Size	.95	1.55
Weighted Mean *d* (95% CI)	.90 (.80 / 1.01)	1.46 (1.07 / 1.22)
Interquartile Range (Q1 – Q3)	.64 – 1.79	1.30 – 2.32
Proportion of Samples Producing Positive Effects	100%	100%
Number of Studies	19	15
Number of Samples	29	22
Number of Participants	1795	1545
Median # Participants / Sample	57	58

Note. CI = confidence interval; Q1 = first quartile (25th percentile); Q3 = third quartile (75th percentile).

One point to note from Table 1 is that 100% of effect sizes were in the expected direction. It is important to remember that the effect size values presented in the current study have been aggregated by sample. That is, a given study might have used the same sample multiple times with different tests or different comparison groups. There would be multiple effects associated with that sample. These effects were denoted by a pooled effect size estimate, calculated using DSTAT. Likewise, one study might have used a given test, or tests that are essentially redundant, with more than one set of samples (i.e., two different sets of malingerer-nonmalingerer pairs). The effect sizes associated with these tests were aggregated into a single effect size, using SPSS. Thus, although individual effect sizes might not have been in the expected direction, when they were pooled or aggregated, the overall effect was in the expected direction.

Table 1 presents the unweighted mean and median effect sizes, as well as *d.* which reflects the mean effect sizes weighted by their sampling errors. The latter value, *d.*, is the value presented in subsequent tables. The values of *d.* indicated that overall, when compared to normal controls, malingerers score about 0.5 of a standard deviation worse than when compared to brain-injured persons. This is a significant difference ($\chi^2 = 40.06$, $p < .0001$). It is important to remember, however, that these values have not yet been screened for modifying variables, such as age, education, and the like. These values will change in subsequent analyses as the subsamples are altered to control for extraneous variables.

Finally, the number of samples reported in Table 1 should be noted. Recall that after exclusion of several studies for methodological and other reasons (given above), there were still 66 samples that included usable data. However, only 29 samples are listed in the Brain-Injured column in Table 1. This reflects the fact that only 29

comparison groups or tests were relevant to that particular analysis. Similar logic applies to the number of studies in the Normal Control column. The impact of this significant attrition will be considered in the Discussion.

Within-Group Effect Size Heterogeneity

After ruling out potential confounds such as outliers, the search for patterns among the effect sizes could begin. The first step in this process involved examination of within-group heterogeneity for the weighted average effect sizes (*d.*). The majority of analyses in this study resulted in significant heterogeneity within groups, indicating, as discussed above, that the samples may not have a common underlying effect. Heterogeneity can be attributed to noise (error), or to systematic variability contributed by a third, moderating variable, such as study characteristics. If heterogeneity results from the latter, it suggests that the groups could be further broken down depending on values of the moderating variable. Moderating variables are often identified through examination of correlations between group ES and values of the moderator (e.g., age, years of education, etc.).

Therefore, correlations were calculated between average unweighted *d*'s and potential moderators, including group mean ages, years of education, percent male, malingerer group head injury severity (where applicable), and time since injury (where applicable). No significant correlations were found for any of these variables and thus they are not presented here. As already noted, the correlation between sample size and effect size was significant, indicating that it was a moderating variable. However, heterogeneity was not significantly decreased when studies were divided into groups based on sample size (*p*'s < .0001).

It is possible that some other third variable that was not reported in the studies systematically affected ES. Another method of discovering moderating variables is to subdivide samples into groups according to values of potential moderators (e.g., presence/absence, mild/moderate/severe, etc.), and then examine the resulting ES (Hedges & Olkin, 1985). This method was used in the present investigation. Few of these efforts were successful in reducing heterogeneity in this sample of studies. In particular, heterogeneity was minimized when groups were divided and re-divided to the point that only small cell sizes remained. Small cell sizes prohibit statistical comparison of data and therefore division of groups to that extent was not employed here.

In summary, attempts to minimize heterogeneity in these analyses were unsuccessful. However, this method did uncover other variables that had a significant impact on effect sizes without significantly decreasing heterogeneity. These variables will be discussed below, where applicable.

Tests of Major Hypotheses

Participant Characteristics as Effect Size Moderators

Brain-Injured Comparison Group

It was anticipated that the different types or groups of malingerers (i.e., Naïve, Coached, Litigants, and Suspected Malingerers) would affect ES magnitude. The effect of Malingerer Group status on neuropsychological test scores is presented in Table 2. The expected pattern of effect sizes (ES) was as follows: Naïve > Coached $\geq$ Litigants $\geq$ Suspected malingerers.

Table 2

Test Score Differences Between Brain-Injured Individuals and Malingering Groups

Malingering group	# of studies	Effect size	H_W	95% CI
Experimental Malingerers	24	.91	146.44*	.80 / 1.01
Coached	8	1.14[a]	39.10*	.94 / 1.35
Naïve	16	.81[a]	100.13*	.68 / .94
Clinical Malingerers	6	.88	14.48	.62 / 1.13
Litigants	3	1.06	+	+
Suspected	3	.76	+	+

Note. CI = confidence interval. * $p < .05$. + = There were too few studies in these groups to allow interpretation. Therefore, the Litigant and Suspected Malingerer groups were combined into a Clinical Malingerer group to permit statistical analysis. The Litigant and Suspected Malingerer effect size values are presented only for interest. [a] = Effect sizes with the same superscript are significantly different ($p < .05$).

Table 2 reveals few significant differences among malingering group effect sizes. The Experimental Malingerer group was comprised of a combination of Coached participants (who were given instructions to avoid detection) and Naïve participants (who were not given such instructions). Experimental Malingerers produced test score differences from brain-injured participants ($d = .91$) that were not significantly larger than those of the Clinical Malingerer participants who were seeking compensation ($d = .88$; $\chi^2 = .04$, *ns*).

However, within experimental participants, information about how avoid detection or how to fake a head injury did have an effect, but in an unexpected direction. It was anticipated that participants who were coached by being given instructions on how to avoid detection or how to fake a head injury would produce scores that were similar to those of actual brain-injured participants, but this was not found. Coached malingerers ($d = 1.14$) produced test scores that were significantly worse than those of Naïve malingerers ($d = .81$; $\chi^2 = 7.21$, $p = .007$).

Demographic characteristics of malingerer groups. As can be seen in Table 3, the groups differed significantly in terms of age ($F = 20.29$, $p < .000$) and education ($F = 47.26$, $p < .000$). Specifically, within the Malingering groups (Group 1 or Experimental group in analyses) the Naïve and Coached participants were younger than Litigants or Suspected malingerers ($p < .001$). Litigants and Suspected malingerers had significantly less education than the Naïve or Coached groups ($p < .001$). Also, the two clinical groups, Litigants and Suspected Malingerers differed in the proportion of head injuries that were Mild, Moderate, and ranging in severity ("Range").

Within the brain-injured comparison group (Group 2 in analyses) there were no differences in terms of age. There were, however, differences in the amount of education, with brain-injured participants who were compared to Naïve malingerers having more education than did brain-injured participants who were compared to Suspected malingerers ($p < .001$). The comparison groups also differed in terms of time since injury ($F = 11.23$, $p < .000$).

In spite of these statistical differences, the demographic differences between the groups did not appear to affect ES: there was no systematic association between these variables and ES as demonstrated by nonsignificant Pearson correlations (not shown). Furthermore, the results of the next set of analyses (proportion of cognitive domains) indicated that attempting to control for any differences by equating groups on demographic variables was not necessary.

The most important factor that varied across type of malingerer was the differing proportions of cognitive domains evaluated in the studies (e.g., Litigants were comprised entirely of Malingering tests). This was viewed as more important because of the expectation that Malingering tests would produce larger ES than tests from other

domains. This could be a confounding variable. Therefore, an attempt was made to control for the potentially confounding effect of different proportion of test domains (see below).

To evaluate this potentially confounding effect of proportion of the samples using tests of a given cognitive domain, the data were re-analyzed by dividing them into two subsets: tests from the Malingering domain, and tests from all other domains except Malingering.

Table 3

Study Characteristics for Malingering Groups Using All Tests

Study Characteristics	Naïve	Coached	Litigants	Suspected
Number of Studies	16	8	3	3
Malingering Group				
Age	27.96 (5.80)[a]	25.99 (7.96)[b,d]	36.13 (2.79)[a,b]	37.62 (.97)[a,d]
Education	14.52 (1.05)[a,b]	13.0 (.89)[b]	11.10 (.65)[a,b]	12.58 (.45)[a]
Percent male	61% (26)	29% (4)	92% (0)	N/a
Head injury severity	N/a	N/a	25% mild 38% mod. 38% range	81% mild 19% range
Comparison Group				
Age	33.59 (5.42)	35.38 (7.96)	37.49 (2.01)	32.18 (.88)
Education	13.12 (.62)[a]	12.58 (1.0)	12.64 (.57)	12.16 (.78)[a]
Percent male	54% (18)	62% (6)	69% (0)	N/a
Time since injury (months)	36.21 (28.04)[a,b]	62.90 (22.34)[a]	N/a	58.29 (16.45)[b]

Note. Mod. = moderate. Values in parentheses represent standard deviations. Values within rows with matching superscripts are significantly different from each other ($p < .005$).

Malingering domain only. When type of cognitive domain was restricted to Malingering tests, group differences in ES disappeared. That is, when only Malingering tests were used, participants given instructions on how to avoid detection (Coached) did not perform differently from malingerers given no such instructions (Naïve) or from

participants seeking compensation (i.e., Clinical Malingerers; overall $\chi^2 = .25$, *ns*). These ES are presented in Table 4.

Table 4

Test Score Differences Between Brain-Injured Individuals and Malingering Groups

Malingering group	# of studies	Effect size	H_W	95% CI
Malingering Tests Only				
Coached + Naïve	13	1.11	71.15*	.95 / 1.27
Coached	5	1.09	38.49*	.84 / 1.34
Naïve	8	1.13	32.57*	.92 / 1.33
Clinical Malingerers	5	1.04	10.02	.78 / 1.31
Litigants	3	1.06	+	+
Suspected	2	1.03	+	+
All Other Domains				
Coached + Naïve	12	.68[a]	63.98*	.53 / .83
Coached	4	1.05[a,b]	7.35	.71 / 1.38
Naïve	9	.63[b]	53.99*	.49 / .79
Suspected	2	.61	+	+

Note. * $p < .05$. + = Too few studies in these groups used tests in given cognitive domain. Where possible, Litigant and Suspected Malingerer groups were combined into a Clinical Malingerer group to permit statistical analysis. The Litigant and Suspected Malingerer effect size values are presented only for interest. Values with matching superscripts within rows are significantly different from each other ($p < .05$).

Although group differences were eliminated when only Malingering tests were considered, all group *d*'s were still significantly different from zero (in the range of 1.03 to 1.13). Confidence intervals were almost identical across the groups, and the confidence intervals did not overlap with zero. This indicates that on Malingering tests, experimental malingerers or people seeking compensation scored approximately one standard deviation worse than actual brain-injured individuals. When tests were restricted to the Malingering domain, Malingerer group differences in proportion of differing severities of head injury were also minimized (see Table 5).

All other domains. When all domains other than Malingering were considered, group differences in effect sizes remained (see Table 4). Because studies evaluating Litigants used only malingering tests, Litigants' scores on tests from domains other than Malingering could not be evaluated. Suspected malingerers performed roughly six-tenths of a standard deviation worse than brain-injured participants, but there were too few studies of this type to permit statistical comparison (see Table 4). Unlike in the Malingering domain, Coached malingerers scored one standard deviation worse than brain-injured participants, but Naïve malingerers scored similarly to Suspected malingerers, only two-thirds of a standard deviation worse than brain-injured individuals. The performances of Coached and Naïve participants were significantly different (χ^2 = 4.92, p = .03).

Coached participants' scores did not differ between Malingering tests and all other domains (χ^2 = .04, *ns*). Naïve participants scored significantly worse on Malingering tests than on tests from any other domain (χ^2 = 14.22, p = .05 post-hoc). Again, there were too few samples of Suspected malingerers using tests from domains other than Malingering to compare their average scores across domains.

Table 5

Study Characteristics for Malingering Groups Using Malingering Tests

Study Characteristics	Naïve-Malingering	Coached-Malingering	Litigants	Suspected-Malingering
Number of Studies	8	5	3	2
Malingering Group				
Age	23.60 (2.99)[a,b]	23.99 (5.74)[c]	36.13 (2.79)[a,c]	36.83 (1.36)[b,c]
Education	14.32 (1.35)[a,b]	13.19 (.66)[a,b]	11.10 (.65)[b,c]	12.78 (.53)[a,c]
Percent male	40% (17)	30% (0)	92% (0)	N/a
Head injury severity	N/a	N/a	25% mild 38% mod. 38% range	40% mild 60% range
Comparison Group				
Age	38.79 (6.79)	35.30 (.94)	37.49 (2.01)	32.64 (.05)
Education	12.87 (.48)	12.72 (.82)	12.64 (.57)	12.90 (1.09)
Percent male	57% (7)	60% (0)	69% (0)	N/a
Time since injury (months)	53.45 (34.09)[a]	76.06 (13.52)[a]	N/a	47.75 (.00)[a]

Note: Mod. = moderate. Values in parentheses represent standard deviations. Values with matching superscripts within rows are significantly different from each other ($p < .05$).

Normal Control Comparison Group

It was anticipated that malingerers' scores on neuropsychological tests would be poorer than would those of brain-damaged individuals. Since brain-damaged individuals usually score worse than normal individuals on many of these tests, it was also expected that malingerers would score below these normal individuals. It was thought that the difference between average scores (i.e., the effect size) obtained by malingerers and normals would be larger than the difference between malingerers and brain-injured participants.

Malingering domain only. Table 6 presents the effect sizes produced by experimental malingerers when compared to normal controls. When analyses were restricted to Malingering tests only, there were no differences between the Coached and Naïve groups ($\chi^2 = .003$, *ns*). In other words, when compared to normal controls,

participants who did not receive instructions on how avoid detection or how to fake a head injury (i.e., Naïve group, $d = 1.28$) did not score significantly differently from participants who received such instructions (Coached group, $d = 1.27$).

Table 6

Test Score Differences Between Normal Healthy Controls and Experimental Malingering Groups

Malingering group	# of studies	Effect size	H_W	95% CI
Malingering Tests Only				
Coached + Naïve	12	1.26	99.41*	1.12 / 1.40
Coached	4	1.27	30.09*	1.03 / 1.51
Naïve	8	1.28	69.30*	1.08 / 1.43
All Other Domains				
Coached + Naïve	14	1.20	46.57*	1.06 / 1.34
Coached	4	1.00[a]	5.82	.77 / 1.24
Naïve	10	1.31[a]	36.95*	1.14 / 1.48

Note. * $p < .05$. Values with matching superscripts within rows are significantly different, $p < .05$.

All other domains. When all cognitive domains except Malingering were used in the analyses, results indicated that malingering group membership had a significant effect on ES. The Naïve group ($d = 1.31$) scored significantly worse than the Coached group ($d = 1.00$; $\chi^2 = 4.34$ p = .04).

There were no differences in effect sizes within malingering groups depending on the cognitive domains considered (Coached Malingering vs. Coached All Other Domains $\chi^2 = 2.40$, *ns* post-hoc; Naïve Malingering vs. Naïve All Other Domains $\chi^2 = .18$, *ns*; see Table 6). Similarly, the performances of the combined malingerer groups were not significantly different when the Malingering domain and all other domains were compared. That is, the ES for Coached + Naïve malingerers using all domains other

than Malingering (d = 1.20) and the ES using only Malingering tests (d = 1.26) were not significantly different (χ^2 = .33, *ns*).

Table 7 displays demographic characteristics for the controls and malingerers, indicating that the groups are comprised of similar participants. Experimental malingerers tended to be older than normal controls (F = 6.79, p = .01). It is unlikely that the observed effect sizes are due solely to some artifact of the group compositions, because ES were not correlated with age, as noted above.

Table 7

Study Characteristics for Normal Healthy Controls and Experimental Malingerers

Study Characteristics	Normal Controls		Experimental Malingerers	
Age	23.34	(1.94)[a]	24.97	(3.94)[a]
Education	13.93	(1.32)	14.04	(1.49)
Percent male	50%	(11%)	46%	(12%)

Note: Values in parentheses represent standard deviations. Experimental malingerers are comprised of Naïve and Coached malingerers combined. Values with the same superscript are significantly different ($p < .05$).

Contrasts Between Comparison Groups

In clinical practice in general, brain-injured individuals' scores on neuropsychological tests are often worse (more pathological) than those of normal individuals. The results from the current meta-analysis presented above show that malingerers' scores are worse than are those of brain-injured persons. Therefore, it was expected that the difference in test scores between malingerers and normals would be larger than that between malingerers and brain-injured participants. The ES associated with the comparisons using normal controls and brain-damaged individuals using only Malingering tests were compared first (see Table 8).

Contrary to expectations, using normal controls as a comparison group did not result in test score differences that were significantly larger than if brain-injured participants' scores were used for the comparison (see below). However, when the comparison groups are subdivided according to cognitive domains and other study characteristics, there are significant differences between the groups (this is discussed below in the section entitled, "Comparison Among Cognitive Domains").

Table 8

Malingering Test Score Differences For Malingerers vs. Normal Controls and Brain-Injured Individuals

Comparison group	# of studies	Effect size	H_W	95% CI
Normal Controls				
Coached + Naïve	12	1.26	99.41*	1.12 / 1.40
Coached	4	1.27	30.09*	1.03 / 1.51
Naïve	8	1.28	69.30*	1.08 / 1.43
Brain-Injured				
Coached + Naïve	13	1.11	71.15*	.95 / 1.27
Coached	5	1.09	38.49*	.84 / 1.34
Naïve	8	1.13	32.57*	.92 / 1.33

Note. * $p < .05$. To minimize extraneous variability, only ES from the Malingering cognitive domain are compared here.

Malingering domain only. Table 8 presents a combination of data from previous tables (Tables 4 and 6) for ease of comparison. It displays that the malingering test score differences between malingerers and normal controls ranged from $d = 1.26$ to 1.28. The differences between malingerers and brain-injured individuals ranged from $d = 1.09$ to 1.13. Although ES of malingerers compared to normal controls tended to be slightly larger than ES of malingerers compared to brain-damaged participants, this difference was not significantly different ($\chi^2 = 1.85$, *ns*).

All other domains. Similar to Table 8, Table 9 presents a combination of data from Tables 4 and 6. The effects presented for comparison in Table 9 are those of the malingerers' differences from normal controls and brain-injured samples, using all cognitive domains other than Malingering. The test score difference between Naïve malingerers and normal controls was $d = 1.31$, whereas the difference between the Naïve group and brain-injured participants was $d = .63$. In contrast to the results using only Malingering tests, the difference between these d's was significant ($\chi^2 = 32.72$, $p < .0001$). The test score difference between Coached malingerers and normals was $d = 1.00$, whereas the difference between the Coached group and brain-injured participants was 1.05. The difference between these d's was not significant ($\chi^2 = .05$, *ns*).

Table 9

Test Score Differences For Malingerers vs. Normal Controls and Malingerers vs. Brain-Injured Individuals Using All Other Cognitive Domains

Comparison group	# of studies	Effect size	H_W	95% CI
Normal Controls				
Coached + Naïve	14	1.20	46.57*	1.06 / 1.34
Coached	4	1.00[a]	5.82	.77 / 1.24
Naïve	10	1.31[a]	36.95*	1.14 / 1.48
Brain-Injured				
Coached + Naïve	12	.68[a]	63.98*	.53 / .83
Coached	4	1.05[a,b]	7.35	.71 / 1.38
Naïve	9	.63[b]	53.99*	.49 / .79

Note. * $p < .05$. Values with matching superscripts within rows are significantly different from each other ($p < .05$).

Impact of Student vs. Normal Community Participants

One variable that was not anticipated to be a moderator of effect sizes was the source of the experimental malingerers. Analyses revealed, however, that effect sizes

depended on the proportion of students and normal participants in the experimental groups. These ES are presented in Table 10.

It was not possible to evaluate the differential effect of the source of experimental malingerers in comparison to normal controls because all studies using such a comparison group used only students as the malingering group; no community participants were included in this group of studies.

Malingering domain only. When normal volunteers were used as the malingering group and only the Malingering domain was considered (as in other analyses), they performed more than one and one-half standard deviations worse than brain-injured non-malingerers ($d = 1.56$). On the other hand, when students were used as the malingerer group, they scored significantly less poorly on malingering tests than did the normals ($d = .90$; $\chi^2 = 23.39$, $p = .0000$).

Table 10

Test Score Differences Between Brain Injured Individuals and Experimental Malingerer Subgroups

Malingering group	# of studies	Effect size	H_W	95% CI
Malingering Tests Only				
Normals	4	1.56	19.50*	1.84 / 1.28
Students	9	.90	36.89*	1.09 / .70
All Other Domains				
Normals	5	.35	20.33*	.16 / .54
Students	7	1.15	16.05*	.92 / 1/38

Note. * $p < .05$

All other domains. In contrast to the results from the Malingering domain, when all cognitive domains other than malingering were considered, normal volunteers scored significantly less poorly than the students ($d_{normals} = .35$; $d_{students} = 1.15$; $\chi^2 = 27.55$, $p <$

.0001). The ES associated with normal volunteers using the Malingering domain and all other domains were significantly different ($\chi^2 = 49.16$, $p < .0001$). However, the students' test scores did not change between the Malingering domain and all other domains ($\chi^2 = 2.78$, *ns* post hoc).

Study Characteristics as Effect Size Moderators

Study Quality

It was expected that ES would vary with the methodological quality of the studies. To reiterate, quality was coded as "methodologically sound" if the study possessed any of the following: random selection or assignment, consecutive referrals, matching on demographic variables, or manipulation check. Quality was coded as "other" if none of these elements were present. The effect of Study Quality (i.e., all studies vs. methodologically sound studies only) on ES is presented in Table 11. Overall, the ES associated with methodologically sound studies was not significantly different from the ES for all studies (including studies that were not as methodologically sound). Sound studies were associated with an average *d* of .90; the ES associated with all studies was .89, a nonsignificant difference ($\chi^2 = .01$, *ns*). This suggested that, contrary to expectations, study quality did not have an impact on the observed test score differences between malingerers and non-malingerers. There were no significant differences in participant demographic characteristics between the samples of studies classified as "Sound" and "Other" (see Appendix C).

Table 11

Test Score Differences Between Brain-Injured Individuals and Malingerers Across Study Quality

Study Quality	# of studies	Effect size	H_T	95% CI
All Studies	29	.89	160.08*	.79 / 1.00
Sound Studies Only	24	.90	145.94*	.79 / 1.01

Note. * $p < .05$.

Type of Compensation

The manner in which experimental malingerers were compensated for participation in studies was found to have an impact on ES. In general, when research participants are drawn from university samples, they are usually offered course credit in return for participation. Community samples are sometimes given money, but in the studies obtained for this investigation, they were usually given no compensation. In malingering research, in addition to course credit or other compensation, malingerers are sometimes told that they will be given extra incentives to try to be believable while they are faking. In fact, all participants usually receive the "extra" incentive. The point is to try to make the artificial situation seem more like when a head-injured person seeks compensation. In such a real-life situation, the person usually knows that if they are too obvious in their exaggeration of symptoms, they will be discovered, and they will not receive damages or other settlement. Similarly, research participants given additional incentives are told that if they are too obvious, they will not receive the extra incentive. Under such a condition, it was hypothesized that participants would avoid exaggeration and would perform more like actual brain-injured individuals. However, this was not found in the present investigation.

Brain-Injured Comparison Group. The effect of type of compensation on neuropsychological test score differences between malingerers and brain-injured non-malingerers is presented in Table 12. When given an additional incentive to avoid detection, malingerers produced test scores that were significantly worse (d = 1.40) than those of malingerers who were given course credit for participation (d = 1.02) or no compensation (d = 1.05; χ^2 = 6.22 and 5.24, $p < .05$, respectively).

Normal Control Comparison Group. The effect of type of compensation on test score differences between malingerers and normal controls is also presented in Table 12. Unfortunately, it was not possible to statistically contrast the effect size associated with the additional incentive condition due to small cell size. The difference between the effects associated with compensation by course credit (d = 1.40) or no compensation (d = 1.46) was not statistically significant (χ^2 = .10, *ns*).

Table 12

Test Score Differences Between Brain-Injured Individuals and Experimental Malingerers Across Incentives

Malingering Incentive	# of studies	Effect size	H_W	95% CI
Brain-Injured Comparison				
Course Credit	8	1.02	22.57*	.80 / 1.23
None/Volunteer	9	1.05	27.25*	.84 / 1.26
Additional Incentive	7	1.40	41.01*	1.19 / 1.61
Normal Control Comparison				
Course Credit	12	1.40	93.37*	1.26 / 1.55
None/Volunteer	4	1.46	1.88	1.17 / 1.75
Additional Incentive	3	1.58	+	+

Note. *$p < .05$. + = There were too few studies in this group to permit statistical comparison. The Additional Incentive effect size is presented only for interest.

Effect sizes for the type of compensation did differ between the brain-injured comparison group and the normal control comparison group. The effect sizes were

significantly different between comparison groups for the course credit ($\chi^2 = 8.49$, $p < .05$) and the volunteer conditions ($\chi^2 = 4.88$, $p < .05$).

Comparison Among Cognitive Domains

The central research question of this meta-analysis was whether different domains of cognitive functioning demonstrated differential ability to distinguish between malingerers and non-malingerers. This was examined comparing malingerers first to brain-injured non-malingerers, and second to normal control non-malingerers. The magnitude of the difference in average effect sizes between these sets of comparisons was then examined.

Brain-Injured Comparison Group

Several cognitive domains were not tested in enough studies using brain-injured individuals as the comparison group to permit analysis – these were the Intellectual, Language, Psychomotor, and Sensory-Perceptual domains. These domains are also typically unrelated to either lay beliefs or evidence from the literature about cognitive sequelae of head injury. Effect sizes for the domains used in four or more studies are presented in Table 13.

Table 13

Test Score Differences Between Malingerers and Brain-Injured Individuals Across Cognitive Domains

Cognitive Domain	# of studies	Effect size	H_W	95% CI
Recognition	4	1.50	33.70*	1.19 / 1.80
Malingering	18	1.10	80.23*	.96 / 1.24
Malingering, reanalysis**	8	1.33	32.19*	1.14 / 1.52
Attention	7	1.07	30.40*	.90 / 1.24
Executive	4	.58	26.78*	.35 / .80
Visuospatial	8	.24	41.90*	.07 / .41
Recall	4	.07	4.21	-.15 / .30

Note. ** Malingering values based on final composition of two student studies. * = $p < .05$.

It was anticipated that the Malingering domain of tests would perform better than other domains in differentiating between malingerers and brain-injured non-malingerers. Consistent with this expectation, malingerers performed more poorly ($d = 1.10$; on average, more than one standard deviation worse) than actual brain-damaged individuals on tests from the Malingering domain. However, malingerers performed even more poorly than brain-injured participants on tasks of Recognition memory ($d = 1.50$). In other words, the scores produced by malingerers on Recognition tests were an average of 1½ standard deviations worse than those produced by brain-injured participants.

Another unexpected finding was that malingerers performed roughly one standard deviation worse than non-malingerers on tests of Attention ($d = 1.07$). A smaller, but still significant, finding was that malingerers also scored over ½ standard deviation worse than non-malingerers on tasks tapping Executive functions ($d = .58$). The finding that malingerers scored approximately ¼ of a standard deviation worse than brain-injured non-malingerers on Visuospatial tasks ($d = .24$) was surprising. Of the

cognitive domains used in enough samples to permit statistical analysis, the Recall domain produced the only result that was nonsignificantly different from zero ($d = .07$; confidence interval overlaps with zero).

Demographic Characteristics. As can be seen in Table 14, the Malingerer group did not differ across cognitive domains on demographic variables such as education ($F = 2.28$, *ns*), head injury severity ($\chi^2 = 6.05$, *ns*), proportion of the sample derived from a given malingering group ($\chi^2 = 11.77$, *ns*), or proportion of the sample that was male ($F = 2.07$, *ns*). The mean age in the Visuospatial domain was significantly greater than the mean age in only one other domain, the Executive domain ($F = 2.61$, $p = .03$). The groups varied somewhat in terms of the proportion of head injuries that were Mild, Moderate, and comprising a Range of severities, but the absolute numbers of studies falling into these categories were small and therefore likely had a minor impact on ES. Approximately 80% of studies in all of the cognitive domains were comprised of studies in which malingering participants did not have a head injury (i.e., experimental malingerers).

The brain-injured comparison group did not differ across cognitive domains on education. The mean age for the Malingering domain was significantly older than the Visuospatial, Attention, and Recognition domains ($F = 6.73$, $p < .001$). The Executive domain was comprised of a significantly greater proportion of males than were the Visuospatial, Attention, and Recognition domains ($F = 6.38$, $p < .003$).

Malingering domain demographics. The Malingering domain in particular included participants who had a longer time since injury, although this was only statistically greater than the length for the Attention group ($F = 3.14$, $p = .02$). No significant differences in ES were found when this factor was controlled (i.e., studies with

Table 14

Study Characteristics for Cognitive Domain Comparisons

	Attention	Executive	Malingering *	Recall	Recognition	Visuospatial
Number of Studies	7	4	8	4	5	8
Malingering Group						
Age	29.67 (6.63)	23.92 (6.93)[a]	32.36 (6.70)	31.60 (5.25)	32.68 (8.19)	32.34 (6.36)[a]
Education	14.34 (.87)	13.71 (.68)	13.18 (1.32)	14.56 (1.11)	13.13 (1.11)	14.11 (1.42)
Percent male	69% (35%)	47% (6%)	30% (0%)	54% (0%)	77% (40%)	62% (33%)
# of Experimental / Clinical Malingerers	6 naive / 1 suspected	2 naïve, 2 coached	3 naïve, 3 coach / 2 suspected	3 naïve / 1 suspected	2 naïve, 1 coach / 1 suspected	6 naïve, 1 coach / 1 suspected
% Student Samples	29 %	75 %	25 %	25 %	20 %	25 %
Head injury severity	94% n/a	88% n/a	21% mild 16% range 63% n/a	20% mild 80% n/a	30% mild 70% n/a	5% mild 95% n/a
Comparison Group						
Age	29.80 (2.76)[a]	32.20 (2.78)	35.26 (2.41)[a,b,c]	32.01 (1.11)	30.72 (3.79)[b]	31.93 (3.38)[c]
Education	13.12 (.57)	12.78 (.66)	12.68 (.99)	13.06 (.95)	12.36 (.82)	12.87 (1.01)
Percent male	47% (18%)[a]	77% (8%)[a,b,c]	60% (0%)	68% (0%)	40% (11%)[b]	48% (15%)[c]
Time since injury (months)	23.33 (12.52)[a]	31.92 (5.78)	64.14 (18.69)[a]	42.81 (33.52)	47.03 (1.24)	35.38 (24.47)

Note: * = Malingering group demographics are based on final sample composition of two student studies. Values in parentheses are standard deviations. Values with same superscript within rows are significantly different from each other ($p < .05$).

long times since injury were randomly dropped[1] until the mean time was 33.34 months, consistent with other domains). Initially, malingerers scored on average over one standard deviation worse than non-malingerers on tests from this domain (d = 1.10). When only studies including comparison groups that had a long time since injury were excluded, malingerers' average effect size did not increase significantly (d = 1.25; χ^2 = 0.79, *ns*). As removing these studies made no significant difference, the studies were retained, and the Malingering with Short Time ES is not reported in Tables 13 or 14.

Most importantly, the Malingering domain also differed notably from other domains in terms of the proportion of participants who were students vs. normal community participants. Most of the other domains were comprised of approximately 20% to 30% students, 50% to 60% normal community participants, and 15% to 25% suspected malingerers. In contrast, the Malingering domain was initially comprised of 50% students, 22% normals, 17% litigants, and 11% suspected malingerers. As already described, the source of the experimental participants had an impact on effect sizes. Therefore, studies involving student samples were randomly dropped to bring the proportion of students in the Malingering domain in line with other domains. Also, Malingering domain studies using Litigants were completely dropped from analyses because no other domains included Litigants.

The final analysis for Malingering included 2 student samples, 4 normals, and 2 suspected malingerers. This resulted in a significant change in ES. When a large

[1] Studies were chosen for removal by placing the study identification number for all student studies into a container and blindly selecting studies to be removed until the proportion of student samples was consistent with that of other cognitive domains.

proportion of the malingering samples were drawn from student populations, malingerers scored on average about one standard deviation worse than non-malingerers ($d = 1.10$). When student samples were randomly dropped out, malingerers' average score on Malingering tests worsened significantly in comparison to non-malingerers ($d = 1.33$; $\chi^2 = 3.70$, $p = .05$). Clearly, as found previously, the source of the experimental participants did have an impact on ES.

The fact that differing proportions of types of participants significantly affected ES indicated the necessity of equalizing the proportion of studies using students and normal participants, in addition to equalizing the proportion of Litigants and Suspected Malingerers. Fortunately, all cognitive domains other than Malingering (which had already been equalized) and Executive already had a roughly equal proportion of participant groups. Analysis then turned to focus on rectifying the unequal proportion of participant groups in the Executive domain.

Equalizing participant groups in the Executive domain. The Executive domain was difficult to equate to other domains in terms of proportion of participant groups because it included no clinical malingerers (see Table 14). All of the studies in this domain that used clinical malingerers had already been dropped from analyses because they did not meet criteria for inclusion (e.g., they used Comparison group participants with mild or severe head injuries). In order to attempt to evaluate the impact of equating the cognitive domain groups on the basis of source of malingerer groups, three studies that did not meet inclusion criteria using clinical malingerers in the Executive domain were added to the analysis. The ES associated with this domain changed from .58 to .48, but this difference was not significant ($\chi^2 = .39$, *ns*). Therefore, the three undesirable clinical malingerer studies were again removed.

Effect sizes for equated cognitive domains. After participant and study characteristics were approximately equated across cognitive domains, effect sizes for the domains could be more accurately compared.

The Recall category was evaluated in enough studies to permit comparisons with other domains, but malingerers' results were not significantly different from non-malingerers (confidence interval overlaps with zero; see Table 13). The largest test score difference between malingerers and brain-injured non-malingerers was associated with Recognition tasks. This was significantly different from all other domains except Malingering (after modification to the Malingering domain in terms of proportion of participant groups as described above).

The general results for the Malingering, Recognition, and Attention tests (i.e., that of all of the cognitive domains, they were associated with the largest test score differences between malingerers and non-malingerers) were as expected. The fact that Recognition tests produced greater differences between brain-injured participants and malingerers than did other tests was not expected.

Normal Control Comparison Group

As found with studies using brain-injured individuals as the comparison group, studies using normal controls as the comparison group did not often use tests from the Intellectual domain. Thus, no analyses could be conducted using this domain. Similarly, the domains of Attention, Executive functioning, and Language abilities were each examined in only three studies. However, the mean effects for each of these domains fell where they were expected, given results from studies using brain-injured participants. Therefore, these mean ES are presented for interest in Table 15, but no

statistical comparisons were performed using these effects. Effect sizes for the domains used in four or more studies are also presented in Table 15.

Table 15

Test Score Differences Between Malingerers and Normal Healthy Controls Across Cognitive Domains

Cognitive Domain	# of studies	Effect size	H_W	95% CI
Attention	3	1.13	+	+
Executive	3	.90	+	+
Language	3	1.03	+	+
Malingering	12	1.42	96.69*	1.28 / 1.57
Psychomotor	6	1.00	14.22*	.82 / 1.19
Recall	11	1.17	55.55*	1.02 / 1.33
Recognition	5	1.56	4.90	1.30 / 1.81
Visuospatial	4	1.03	1.03	.73 / 1.34

Note. * = $p < .05$. + = Too few studies used tests from these domains to permit comparison; effect sizes are presented only for interest.

It was anticipated that the Malingering domain would perform better than other domains in differentiating between malingerers and normal controls. As expected, malingerers performed more poorly ($d = 1.42$) than normal controls on tests from the Malingering domain. However, as found when brain-injured individuals were used as the comparison group, malingerers also performed poorly on Recognition tasks ($d = 1.56$).

Malingerers scored on average more than one standard deviation worse than normal controls on tests of Recall memory ($d = 1.17$), Visuospatial abilities ($d = 1.03$), and Psychomotor abilities ($d = 1.00$). These results were unexpected and are noticeably different from those obtained when brain-injured individuals were used as the comparison group (when domains were used in enough studies with brain-injured comparison groups to permit analysis). To summarize, malingerers performed at least

one standard deviation worse than did normal controls on tests tapping a variety of cognitive domains.

Contrasts Between Comparison Groups

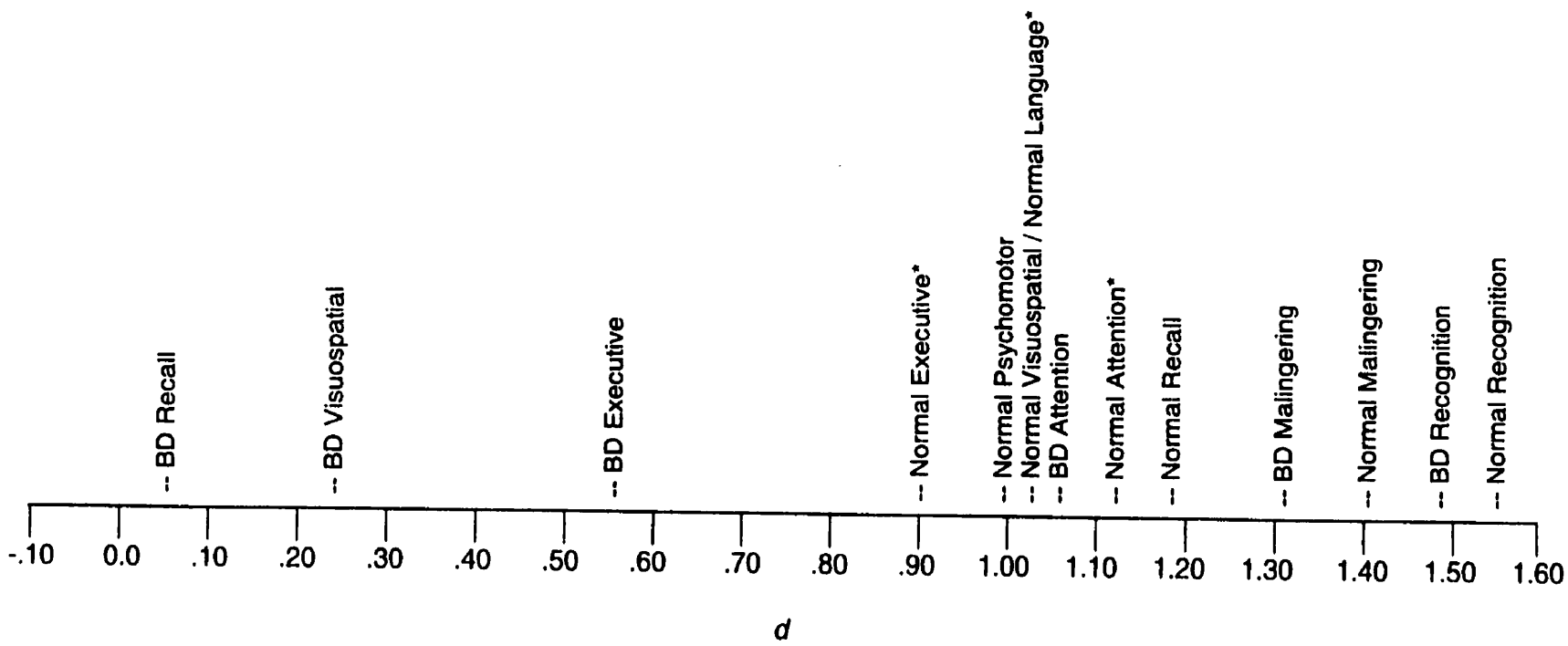
Figure 4 displays the dispersion of effect sizes for the cognitive domains by comparison group. Table 16 presents the statistical contrasts among the cognitive domains according to comparison group. Inspection of Figure 4 reveals that the domains are distributed in roughly the same pattern across the two comparison groups. That is, for both comparison groups, the Recognition domain was associated with the largest effect sizes, followed by the Malingering, Attention, and Executive domains[2]. It is also noteworthy that within this pattern, for the domains used frequently enough to permit comparison, the effects associated with the normal controls tended to be (nonsignificantly) larger than those associated with the brain-injured comparison group (see Table 16). It was expected that effects associated with normal controls would be larger than those associated with brain-injured participants, although it was anticipated that this finding would be significant for all domains.

The ES associated with the Recall domain differed significantly between the normal control and brain-injured comparison groups ($\chi^2 = 61.60$, $p < .001$). The difference between these groups was unexpected because it was not anticipated that the ES associated with the brain-injured participants would be small. It was also not

[2] Recall that there were too few studies of normal controls that used tests from the Attention and Executive domains to permit statistical comparison with other domains. However, given that the pattern of effects associated with these cognitive domains is similar to that found with the brain-injured comparison group, it was thought that these effects were reliable even with small sample sizes.

Figure 5

Distribution of Cognitive Domain Effect Sizes



Note: *These domains were not used in enough studies to permit statistical comparison and therefore their effect sizes might not be reliable; they are placed in this figure only for interest. BD = brain-injured comparison group.

Table 16

Contrast Matrix for Cognitive Domains and Comparison Groups

	Normal Malingering	Normal Psychomotor	Normal Recall	Normal Recognition	Normal Visuospatial	BD Attention	BD Executive	BD Malingering	BD Recall	BD Visuospatial
Normal Malingering										
Normal Psychomotor	12.01*									
Normal Recall	5.31*	1.87								
Normal Recognition	0.81	11.65*	6.31*							
Normal Visuospatial	5.35*	.02	.69	6.87*						
BD Attention	9.66*	.24	.82	9.70*	.04					
BD Executive	38.66*	8.24*	18.49*	31.81*	5.73*	11.62*				
BD Malingering	0.52	5.78*	1.61	1.87	2.78	4.11*	25.20*			
BD Recall	96.43*	38.40*	61.60*	71.90*	25.13*	46.84*	9.54*	68.71*		
BD Visuospatial	106.11*	34.67*	62.50*	69.64*	20.35*	44.51*	5.46*	68.70*	1.29	
BD Recognition	0.18	7.29*	3.43	.09	4.57*	5.81*	22.76*	.78	53.92*	49.60*

Note: Values in cells represent χ^2. BD = brain-injured comparison group. * $p < .05$.

expected that the ES associated with the Visuospatial domain would differ significantly between the two comparison groups ($\chi^2 = 20.35$, $p < .001$).

In general, on neuropsychological tests malingerers performed at least one standard deviation worse than did normal controls, across all types of tests (i.e., cognitive domains). In contrast, malingerers' performance relative to brain-injured controls varied significantly by cognitive domain. The differing amount of inter-domain variability by comparison group was not expected.

DISCUSSION

Overview

This meta-analytic investigation evaluated the degree to which neuropsychological instruments were able to discriminate between head injury malingerers and non-malingerers. Lees-Haley, Smith, Williams, and Dunn (1996) surveyed the frequency of use of various psychological tests by forensic neuropsychologists in personal injury cases. Assuming that malingering is present with some frequency in such cases (Cullum et al., 1991), it is surprising that only one test designed to assess malingering, the Rey 15-Item test, was cited as being used at all (and in only 8% of cases at that) in these evaluations. Given the concerns regarding the usefulness of the Rey 15-Item in detecting malingering (Rogers et al., 1993), it is dismaying that it was the only malingering instrument found to be in use.

Of the other malingering tests in use, Larrabee (1990) concluded that symptom validity testing is time consuming, that it provides little information regarding test performance, and that high-functioning potential malingerers might grow suspicious of the method. Nevertheless, researchers who examine tests designed specifically to assess malingering have continued to focus on the symptom validity paradigm to the relative exclusion of other methods. Rogers and colleagues (1993) described six methods for identifying malingering, including the symptom validity paradigm and performance curve strategies. They noted that the sensitivity of the symptom validity paradigm was very low. They concluded that evaluating an examinee's performance curve showed the greatest promise for accurately classifying malingerers, but that few

studies had researched the performance curve approach. Nies and Sweet conducted an exhaustive review of the malingering literature and concluded that, "50 years of research has not resulted in a general conclusion or consensus regarding detection of malingering" (p.540). The current meta-analysis addressed this crucial issue.

Heterogeneity Among Studies

Prior to examining the central research questions, the degree of heterogeneity that was present among the studies that were sampled needed to be assessed. Rosenthal (1995) pointed out that when significant heterogeneity is present in meta-analytic results, it "alert[s] the meta-analyst to the likelihood that all the effect sizes are not cut from the same cloth and that he or she should try to find the moderator variables accounting for the significant heterogeneity" (p.188). That significant heterogeneity was found in the current investigation suggests that moderator variables were present in the data.

Identification of these moderator variables would suggest how to divide the data to eliminate the heterogeneity. For example, a set of WAIS-III IQ scores might be associated with a significant degree of heterogeneity. Examination of the characteristics of the participants or studies that contributed to this set of scores might reveal that the participants included a subgroup of persons who spoke English as a second language. Once this group was analyzed separately, assuming that the remainder of the participants were homogeneous, the heterogeneity test would be nonsignificant. In this study, the obvious potential moderators, such as age, years of education, proportion of the sample that was male, study quality, and so on were examined. Where applicable, moderator effects were analyzed by computing Pearson correlations or chi-squares

between the value of the potential moderator and the obtained effect sizes (Rosenthal, 1995).

None of the correlational analyses were significant, with the exception of the association between sample size and effect size. When studies were divided into groups using the median *n* as the dividing point, small *n* studies were found to have significantly larger effect sizes than large *n* studies. However, even the effect sizes associated with the larger *n* studies were in the large range (according to Cohen, 1988), which lends further support to the reliability of the magnitude of the effect sizes discussed below. Unfortunately, dividing the studies by sample size did not result in significantly reduced heterogeneity.

Another approach to identifying moderators was to divide the data on the basis of a potential moderator in the absence of significant correlations and run the effect size analyses to see if the heterogeneity was eliminated (Hedges & Olkin, 1985). This approach worked to some extent, but many analyses were still significantly heterogeneous. The only way to fully eliminate heterogeneity was to divide and sub-divide the data until prohibitively small cell sizes remained. And, as Rosenthal (1995) noted, the significance of the heterogeneity is related to the sample size. Therefore, it is not clear whether the heterogeneity was truly eliminated by using small cell sizes, or whether the significance test was influenced by the cell size. In summary, a significant degree of heterogeneity was associated with the analyses in this study, indicating that moderator variables had an effect on the results.

That it was not possible to identify moderators that completely eliminated heterogeneity suggests that either the potential moderators were not reported by researchers, the moderators were not coded by the present investigator, or noise

(measurement error) was present in a large proportion of the studies sampled. The significant heterogeneity limits confidence in the conclusions, since the observed results do not fully reflect the differences among the studies. It cannot be established with absolute certainty that the observed test score differences (effect sizes) were due to the variables that were examined. It is possible that some third variable(s) account for the observed results. This inability to control extraneous variability, or dependence on the quality and methods of studies that have been conducted, is a shortcoming of meta-analysis in general.

Nevertheless, study quality (as assessed in a global form here) did not have an impact on the observed test score differences between malingerers and non-malingerers. Of course, some studies of the poorest quality were excluded prior to analysis because of missing data or lack of a control group. This probably restricted the effect of study quality on test score differences. It appears that if anything, the quality of the research domain as a whole, rather than quality of individual studies, might have had a small effect on the results of this meta-analysis.

Another possible explanation for the significant degree of heterogeneity relates to the model used to analyze the data. This study used a random effects model in that not all possible levels of potential moderator variables were included. The statistical analysis package DSTAT (Johnson, 1989) assumed a fixed effects model for the data, whereby all levels of potential moderator variables were assumed to be included. Therefore, the significant heterogeneity of observed effect sizes is partly an artifact of the disagreement between the models employed and the resulting degrees of freedom.

Summary of Important Findings

As stated above, to date there has been no consensus in the literature regarding the detection of malingering. Of late, researchers appear to have placed emphasis on the symptom validity approach to identification of malingering, but the question of whether neuropsychological instruments can detect malingering has been ignored. In fact, the issue of whether any type of psychological test can reliably identify malingering at all has not been answered. This study addressed these issues, producing some surprising results. The most important finding in this study was that malingerers' performance on tests of recognition memory and malingering fell well below that of non-malingerers. In fact, on tests of recognition memory, the average performance of malingerers was one and one-half standard deviations worse than that of non-malingerers, either normal controls or brain-injured persons. In describing the magnitude of various effect sizes, Cohen (1992) reported that a *d* of .50 is a medium effect size, that is, visible to the naked eye. He described a *d* of .80 as a large effect. The vast majority of effects observed in the present meta-analysis were greater than *d* = 1.00. According to Cohen's descriptors, then, the degree to which neuropsychological instruments were able to distinguish between malingerers and non-malingerers was very large.

Given the large effect sizes and statistically significant findings of this meta-analysis, we should expect that the studies evaluated here included a sufficient number of participants to detect the observed differences between malingerers and non-malingerers. To detect an effect of *d* = 1.00 at the .05 level of significance with Cohen's suggested conventional power level of .80, the study sample should include a total of 17 participants (Cohen, 1988, Table 2.4.1, p. 55). According to Table 2 (above), the

median number of participants per study evaluated here was 57 or 58. Indeed, there were only two samples out of the entire set of 85 samples that included fewer than 17 participants (Millis & Kler, 1995; Prigatano et al., 1997). We can be sure that the vast majority of studies had enough power to detect a malingering effect.

The following paragraphs will discuss the pattern of test score differences (effect sizes) produced by malingerers, and the variables that are associated with the effects. The impressive effect size results for the different cognitive domains will be discussed first, followed by the complex differences in effect sizes among the malingering participant groups. The effect size differences between malingerers and brain-injured and normal controls will end the discussion of the results.

Due to a large amount of attrition in the number of studies available for analysis, it was not possible to test the third hypothesis that particular tests would be better than would others at discriminating between malingerers and non-malingerers.

Explanation of Results and Incorporation with Existing Theory and Research

Cognitive Domain Results

The goal of this meta-analysis was to address the lack of consensus in the literature regarding which tests or methods are most effective at detecting malingering. Initially, it was hoped that enough data would be available to analyze the issue at the level of individual tests, so that a conclusion could be drawn about which test(s) were best to use when evaluating a possible malingerer. Unfortunately, due to a large amount of attrition in the number of studies that met criteria for inclusion, it was not possible to investigate individual instruments. It was thought that investigating the degree to which

different domains of cognitive functioning could distinguish between malingerers and non-malingerers would provide similar information to the individual test analysis. Following Leonberger and colleagues (1992) and Larrabee and colleagues (1985; Larrabee & Curtis, 1995), tests were divided into domains of cognitive functioning. Lezak's (1995) test descriptions provided additional assistance for classifying the tests.

It was expected that the Malingering domain would distinguish between the groups more clearly than other domains. Consistent with this hypothesis, malingerers scored between 1.3 and 1.4 standard deviations worse than either brain-injured or normal control non-malingerers, respectively on malingering tests. A difference of this size means that at most, only 34% of the malingerer and nonmalingerer distributions overlap (Cohen, 1988).

A surprising finding was that malingerers did even worse on tests of Recognition memory. The difference between malingerers and non-malingerers on Recognition tasks was approximately 1½ standard deviations, a difference that would be clearly visible to the naked eye. This indicates that there is only about 29% overlap between the malingerer and nonmalingerer distributions (Cohen, 1988). Malingerers also did significantly worse than either normal or brain-injured non-malingerers on tasks of Attention, Executive, and Visuospatial functioning.

The findings with regard to the differences among the cognitive domains of Malingering and Recognition were more or less as expected. It was expected that Recognition tasks would discriminate well between malingerers and non-malingerers because the performance of brain-injured individuals on these tasks is typically similar to that of normal individuals. In contrast, malingerers typically perform poorly on recognition tests (Brandt, 1992; Iverson et al., 1991). Because many of the Malingering

tests use a recognition paradigm as the basis for the task, it was expected that Malingering tests would also discriminate well between malingerers and non-malingerers.

Since it has been repeatedly demonstrated in the literature that malingerers perform poorly on recognition tasks relative to either brain-injured or normal controls (e.g., Brandt, 1992; Iverson et al., 1991), it was expected that recognition tasks would discriminate between malingerers and non-malingerers. However, it was not anticipated that recognition tests would discriminate better than malingering tasks, because tasks in both the domains of recognition and malingering (at least the symptom validity tests, which made up the majority of tests in this domain) are based on the same principle of recognition abilities. For example, tests used in the recognition category included the RAVLT or CVLT recognition section and the Recognition Memory Test (Warrington, 1984). The Recognition Memory Test is an interesting instrument, because although it was not designed as a malingering test, it possesses many of the qualities of instruments in that category. As its name suggests, it uses a recognition paradigm. It involves presentation of a large number of stimuli that the participant is later asked to recognize from a series of two options. In doing so, it permits calculation of chance or near-chance levels of performance, which would suggest questionable motivation. Thus, the Recognition Memory Test can also be viewed as falling in the symptom validity category of malingering instruments. The hypothesis was that these two groups of tests were essentially equivalent. The results of this study revealed that this is not necessarily the case.

That recognition and malingering tasks are similar could account for the fact that malingerers' performance on recognition tasks was as poor as their performance on

malingering tasks. However, it does not account for why they did significantly worse on recognition tasks (i.e., larger effect size) than on malingering tasks. One potential reason is that malingering tests of the symptom validity type usually present strings of digits or other nonsense data as the stimulus. Recognition tests, on the other hand, present words or pictures of faces. It might be that the use of words or faces instead of numbers somehow facilitates the malingering effect.

In contrast to the expected and observed results for the Malingering and Recognition domains, the results for the domains of Attention, Executive processes, and Visuospatial processes were not expected. The expectations around which domains would be better at differentiating between malingerers and non-malingerers were based on research into the beliefs of laypersons regarding cognitive functioning following head injury (Aubrey et al., 1989; Gouvier, Prestholdt, & Warner, 1988). They were also based on neuropsychological studies of cognitive deficits after head injury (Barth, Macciocchi, Giordani, Rimel, Jane, & Boll, 1983; Dikmen, Machamer, Winn, & Temkin, 1995; Rimel, Giordani, Barth, Boll, & Jane, 1981).

Laypeople tend to believe that memory problems are common following head injury, which is true, but they also tend to hold a variety of misconceptions regarding sequelae of head injury that presumably influence their malingering performance (Aubrey et al., 1989; Gouvier et al., 1988). Impaired performance usually occurs on recall tasks, not recognition. This is likely why for recall tasks, no differences were found between malingerers and brain-injured non-malingerers, while significant differences were found between malingerers and normal controls. In contrast, differences between malingerers and both groups of non-malingerers were significant for recognition tasks.

These findings are consistent with previous research (Brandt, 1992; Iverson et al., 1991).

The findings with regard to visuospatial, executive, and attentional tasks were not anticipated because the cognitive abilities involved with these tests are not in line with common lay beliefs regarding specific deficits following head injury. However, laypeople may tend to believe that cognitive functioning is globally disrupted after head injury (Aubrey et al., 1989; Gouvier et al., 1988). Perhaps, then, malingerers thought that they should do poorly on all tasks, including tasks involving drawing and manipulation of blocks or puzzles. From this perspective, these results are more understandable.

These results are also understandable in terms of neuropsychological research into cognitive sequelae of head injury, which indicates that in addition to memory problems, executive and attention abilities can be significantly disrupted following even a mild head injury (Barth et al., 1983; Rimel et al., 1981). Using essentially similar classifications of cognitive domains as those employed here, Binder and associates (1997) found that the domains of Attention and Memory Acquisition were associated with the largest relative differences between normals and individuals who had suffered a mild head injury. This relative difference in ES among cognitive domains is similar to that found in the current meta-analysis. It seems, then, that although laypersons do not typically list executive and attentional processes as among those affected by head injury, in the present investigation they suppressed performance in these domains in a manner consistent with findings in the literature.

The fact that the observed effect sizes were so large is surprising in and of itself. As noted above, there has been no consensus in the literature on malingering of neurocognitive deficits with regard to which, if any, neuropsychological instrument(s) are

most useful in identification of malingering. The large effect sizes found in this meta-analysis suggest that malingering should be easily observable to the naked eye (Cohen, 1988). It is not clear why the large effects observed here have not frequently been detected in individual studies. It may be, as Schmidt (1996) has noted, that the reliance on significance testing has obscured clinically important results. However, given that large effects are usually found to be statistically significant with even small sample sizes, this explanation may not be applicable. That the studies examined here had unusually large effects (i.e., artifactual sampling problems) is unlikely given that outliers were removed, and that publication bias has been ruled out. The nature of the brain-injured comparison group (i.e., that only individuals with moderate head injuries were included) is an unlikely explanation given that there were few differences in effect sizes between the brain-injured group and the normal controls.

In summary, the results of this meta-analysis indicated that on either Malingering tests or Recognition tasks, malingerers will likely perform significantly worse than either normal controls or actual brain-injured individuals. This information could be helpful (in a context of collateral information about the individual's behaviour and symptoms) to clinicians, judges, and juries in personal injury cases. The information provided by tests of Attention, Executive functioning, and Visuospatial processes could be useful, but such information would less clearly distinguish between malingerers and non-malingerers than would Malingering or Recognition tests.

Differences Among Malingering Participant Groups

The majority of malingering studies have used "experimental malingerers," non-head-injured students or community volunteers who are asked to simulate malingering.

The generalizability of results based on normal volunteers to actual malingerers in clinical settings has been strongly criticized (Nies & Sweet, 1994; Rogers & Cavanaugh, 1983). In an attempt to avoid this generalizability problem, some researchers have identified groups of head-injured individuals who are at high risk for exaggerating or malingering their impairment (Binder, 1993; Binder & Willis, 1991; Greiffenstein et al., 1995). Individuals in litigation for damages following a head injury, and especially those who are suspected to be malingering by clinicians not involved with the research, have comprised the "clinical malingerer" group. Using these "clinical malingerers" has been criticized on the basis that not all litigants are necessarily malingering, and malingerers who have been caught may not be the same as malingerers who have not been caught (Faust & Ackley, 1998). The present meta-analysis compared the average performances of all four groups in order to address the issue of which group(s) should be used in malingering research.

A main hypothesis of this study was that there would be an interaction between cognitive domains and types of participants in the degree to which the tests distinguished between malingerers and non-malingerers. This meta-analysis found a complex interaction between these variables.

One major variable affecting test score differences between groups was the proportion of the tests used that were derived from the Malingering domain. When malingering groups were compared to brain-injured participants, there was significant variability in test score differences across cognitive domains, although not in the expected pattern (i.e., the Naïve group had the largest difference, followed by the other groups). However, when only Malingering tests were considered, the differences between malingering groups were eliminated. This effect of cognitive domain was not

anticipated. It had been expected that differences between malingering groups would hold regardless of cognitive domain. Malingerers on the whole scored more than one standard deviation worse than brain-injured non-malingerers on Malingering tests. A difference of this size indicates that there was only 41% overlap between the distributions of the malingerers and brain-injured non-malingerers (see Cohen, 1988, Table 2.2.1, p. 22).

Overall Results

When extraneous variability was controlled to the greatest extent possible, such as when only Malingering tests were examined, the performances of the different malingerer groups was essentially the same. However, when all cognitive domains other than Malingering were used, there were some differences in average effect sizes between groups. Specifically, Coached malingerers scored significantly worse on neuropsychological tests than did Naïve malingerers, although both groups performed significantly worse than either brain-injured individuals or normal controls. (There were too few samples of Litigants and Suspected malingerers to permit comparison of these individual groups to other groups.)

One potential reason for this lack of variability among groups when considering only malingering tests might be that tests from domains other than Malingering might be more cognitively complicated. For instance, the Wisconsin Card Sorting Test involves having the examinee sort two sets of 64 cards with a variety of coloured symbols according to principles that shift periodically. The principles have to be deduced from the examiner informing the examinee only that he or she is correct or incorrect. In contrast, the symptom validity tests simply require the examinee to determine which of

two stimuli he or she saw/heard/felt in the initial set of target stimuli. The difference in difficulty between the malingering tests and tests from other domains might allow any between-group differences to be expressed.

Another potential explanation for the lack of variability among groups on malingering tests is that tests from domains other than malingering might be sensitive to the different types of malingerers (i.e., students vs. normal community volunteers; Litigants vs. naïve student malingerers). In contrast, Malingering tests are simpler and might be sensitive simply to the presence (vs. absence) of malingering of any type. Thus, on Malingering tests, any type of malingerer (coached, naïve, litigant, or suspected malingerer) would score poorly relative to both normal controls and brain-injured individuals.

Another possibility is that examining only one cognitive domain at a time would eliminate effect size differences between groups. For example, looking only at the Attention domain might eliminate effect size differences between groups in the same way that looking only at the Malingering domain removed differences between groups. In other words, the apparent differences between groups might have been related to a third variable: the proportion of different cognitive domains that were used in the analysis. It was not possible to test this hypothesis because studies investigating Litigants used only malingering tests. Thus, any investigation of individual domains other than Malingering would have been confounded by a lack of Litigants.

Students vs. Community Volunteers. One interesting finding was not hypothesized and only became apparent during data analysis. That was the finding that the source of the experimental malingerers (i.e., whether they were students or community volunteers) had an impact on test score differences. Students, on average,

scored about one standard deviation worse than brain-injured non-malingerers when all cognitive domains were used. Normal volunteers performed better than the students (i.e., more like brain-injured participants), but still significantly worse than the brain-injured individuals. When only the Malingering domain was considered, the difference between the groups was eliminated, so that both groups scored about one standard deviation worse than non-malingerers.

Coached Group Results

It was anticipated that Coached malingerers would perform less poorly on tests (i.e., they would produce smaller effect sizes) than would the Naïve malingerers. The Coached group was given tips on how to avoid detection as a malingerer, or on how to fake head injury or memory impairment. Research has indicated that Coached participants are generally better than Naïve malingerers at avoiding detection (Johnson & Lesniak-Karpiak, 1997; Martin et al., 1993; Rose et al., 1998). This finding has been considered to be so robust that some prominent researchers have urged psychologists to refrain from warning examinees that efforts to malinger would be detected (Youngjohn, Lees-Haley, & Binder, 1999). Contrary to expectations, the present meta-analytic investigation showed that Coached malingerers produced larger effect sizes than did Naïve malingerers; that is, the Coached group performed worse on neuropsychological tests than did Naïve malingerers.

That the Coached malingerers in this study produced results that were so different from expectations needs further explanation. In this study, coaching was analyzed as present vs. absent, because it was hypothesized that different types of coaching would have similar effects. That is, it was thought that coaching participants

on how to avoid detection and coaching on how to fake head injury would have the same effect. However, it is possible that this is not the case. In fact, a study that investigated malingering on the MMPI-2 addressed this issue (Lamb, Berry, Wetter, & Baer, 1994). Lamb and colleagues divided malingering participants into three groups: the first received information on the symptoms of head injury; the second received information on the validity scales of the MMPI-2; and the third received information on both head injury and the validity scales. The results indicated that information on the validity scales produced test scores that were similar to non-malingerers, whereas information on head injury produced test scores that were in the more pathological direction (i.e., larger effect sizes).

If the results of the Lamb et al. study are generalizable to the present investigation, it could be that combining groups who received information on head injury with those who were told how to avoid detection obscured any differential effect that these two instruction sets might have had. The fact that the Coached group as a whole produced a relatively large average effect size suggests that the majority of Coached participants in the present meta-analysis received information on head injury.

Coached malingerers vs. Litigants. Several studies have attempted to examine the effect of coaching on test performance (e.g., Coleman et al., 1998; Hiscock et al., 1994; Lamb et al., 1994; Martin et al., 1993; Rose et al., 1995). These studies are potentially confounded, though, by important differences between experimental coached participants and lawyer-coached clinical litigants. For example, it could be that the thoroughness of the preparation that the experimental coached participants receive is limited in comparison to the coaching a litigating individual might receive from a dedicated lawyer. If this is true, one might expect that coached participants would not

have learned as many skills for concealing their malingering as the lawyer-coached litigants. They would then produce relatively large effect sizes. In contrast, litigants who might be prepared more thoroughly by an attorney would be better able to conceal their malingering, and thus produce smaller effect sizes than coached participants. This hypothesis is consistent with the findings of the present study.

Alternatively, the reality of the situation for litigants (i.e., the possibility of not obtaining a large settlement if malingering is detected) might have played a role in their malingering not being as obvious. Experimental coached participants might not have felt the same pressure to conceal their malingering, since a settlement was not in jeopardy if they were detected. It is likely that, as Nies and Sweet (1994) point out, "motivation to participate in a study, even when compensated financially, is different from the motivation to gain financial reward through litigation" (p. 510). Thus, the Coached malingerers were not concerned about hiding their malingering and scored more poorly than did the Litigants relative to brain-injured controls.

Litigant Group Results

Researchers often identify people who are involved in litigation or who are seeking compensation for injuries as being at increased risk for exaggerating or malingering. The major shortcoming of labelling these litigants as malingerers is that it is thought that simply pursuing compensation is insufficient grounds for labelling a participant as a malingerer; not all of those who seek compensation are necessarily malingerers (Nies & Sweet, 1994). This approach is also unhelpful to clinicians working in workers' compensation or insurance settings, who are asked to identify malingerers

from a sample of Litigants. The validity of labelling Litigants as malingerers was therefore examined in this study.

The present meta-analysis compared Litigants to experimental malingering groups (i.e., experimental coached and naïve malingerers). Haines and Norris (1995) suggested that experimental malingerers are relatively more validly identified in that researchers instruct normal volunteers to act as though they are malingering. The results of the present study indicated that Litigants performed no differently from the experimental malingering groups. However, this result was confounded by the fact that studies involving Litigants used only Malingering domain tests. Comparisons between the Malingering domain and all other domains using Naïve, Coached, or Suspected malingerers showed that use of only Malingering tests obscured effect size differences between groups. Litigants could only be examined using Malingering tests that might have eliminated differences between Litigants and other groups. Therefore, it might be that the results of the current investigation suggesting that Litigants perform similarly to other malingerers might not be accurate.

However, whether or not it is true that Litigants are the same as other malingerers, the results of the present study indicated that it *is* true that Litigants are different from non-malingerers. In this study, Litigants performed significantly differently from brain-injured and normal non-malingerers. Thus, these results suggest that labelling litigants as malingerers is not an entirely invalid research technique.

However, this study examined *groups* of litigants, not individuals. Therefore, the results of this study do not indicate that all litigants should be automatically labelled as malingerers during clinical evaluations of individuals. In a clinical evaluation, diagnosing

an individual litigant as a malingerer should only be considered after carefully weighing all available test and collateral information.

Malingerers Offered Additional Incentives

It was anticipated that the type of compensation offered to experimental malingerers for participation might affect their performance. The results of this study suggest that this is true, but not in a manner that would be expected. Outside of the laboratory, it is presumed that people malinger in order to obtain some desired benefit, often money (American Psychiatric Association, 1994). They know that if they are caught (generally if their malingering is too obvious), they will not receive the benefits they desire. Malingering researchers offer extra incentives to participants, supposedly to make the experimental situation similar to real-life litigation or compensation-seeking situations. Researchers tell participants that if they can successfully fool the examiner into thinking that they actually do have deficits, the participant will receive the extra incentive. Participants are sometimes told explicitly that if they are too obvious, the examiner will think that they are malingering, and they will not receive the incentive. Theoretically, then, malingerers should avoid exaggerating their malingered deficits.

The results from this study are exactly the opposite of this expectation. When offered additional incentives, participants performed more poorly than if they were not given the incentives, regardless of the cognitive domain tested.

Research demonstrates that 10 to 20% of participants are unwilling or unable to malinger, even when given incentives (Pankratz & Binder, 1997). The current investigation revealed that although some percentage of experimental participants might not have been able to malinger (this was not investigated in this study), some of the

experimental participants were quite able to malinger, especially when given incentives. Not only were they able to malinger, but they suppressed their scores to a greater extent than malingerers who were not given such incentives. These results suggest that the offer of additional incentives to successfully fool the examiner induced participants to go overboard in their efforts to malinger, producing scores that were much poorer than non-malingerers. This is contrary to expectations because if, in fact, a "real" malingerer produced scores as poor as this group, he or she would likely have been caught because the scores are so different from those of actual brain-injured individuals.

These results suggest that the practice of offering additional incentives to participants who have been asked to malinger is not warranted.

Which Malingerer Group(s) Should Be Used?

Rogers and colleagues (1993) recommended that researchers should avoid using one type of malingerer in studies because the equivalence of the groups had not yet been established. Furthermore, they noted that there were weaknesses associated with all of the potential malingerer groups. They suggested that malingering research methods should be based on convergent results from both the experimental malingerer and clinical malingerer paradigms. Thus, testing of malingering detection procedures would occur under controlled conditions using experimental malingerers, and the generalizability of these procedures to clinical practice could be evaluated using clinical malingerers. To date, however, the convergence of these methods has not been achieved using traditional experimental design and analysis.

A meta-analysis can achieve a type of convergence of results. The results of the present meta-analysis indicated that if brain-injured individuals were to be used as the

comparison group, all malingering groups would likely perform significantly worse than the brain-injured group. The results of the Lamb et al. (1994) study suggest that the Coached group might exaggerate their performance more (i.e., perform more poorly than) than other malingering groups if they are given information about symptoms of brain injury. In contrast, Coached subjects might perform less poorly than other malingering groups if they are given information about how to elude detection. Any or all groups might be appropriate for studies using Malingering tests, depending on the questions to be addressed. It will be important for researchers to recall that using only Malingering tests appears to eliminate differences among malingerer groups.

If normal controls are to be used as the comparison group, both Coached and Naïve malingerers could be used because in the current meta-analysis, scores produced by both groups were significantly worse than those of normals. However, the Naïve participants might exaggerate their poor performance more than the Coached group. As with the brain-injured comparison group, the type of instructions given to the Coached group will likely affect the magnitude of their poor performance. It was not possible to determine how Suspected malingerers or Litigants would perform relative to normal controls. Generalizing from the results that used brain-injured individuals as the comparison group, though, it is likely that Suspected malingerers and Litigants would perform significantly worse than normal controls.

Of course, it is not possible to achieve an ultimate standard for diagnosing malingering, because many malingerers (likely those that are best at malingering) are never identified. In the current study, Suspected malingerers were those who were unable to fool examiners into believing that they had suffered real deficits. It would be ideal if a group of actual malingerers who were identified after having fooled

sophisticated examiners could be detected and used as the ultimate comparison group. However, we still would not know the characteristics or tactics of malingerers who had not been caught. Therefore, using actual malingerers who have been caught would still be unable to overcome the argument that caught malingerers might differ systematically from those who have not been caught (Faust & Ackley, 1998). Malingering research will likely always have to face the weakness of questionable external validity.

Differences Between Comparison Groups

Researchers in the area of malingering have criticized the practice of comparing malingerers to normal controls (Faust & Ackley, 1998). They cite the fact that in a clinical setting, neuropsychologists never need to distinguish between malingerers and normal individuals (Cullum et al., 1991); rather, they need to distinguish between malingerers and brain-injured non-malingerers. However, the difficulty with relying on brain-injured samples as a comparison group is that not all brain-injured individuals within the group are the same. Generally brain-injured groups have suffered a range of severities of head injury to a variety of locations within the brain, and are tested at a range of times from the injury. Thus, the cognitive performances of these individuals will vary within the group. This limits the utility of brain-injured subjects as a comparison group.

Furthermore, brain-injured persons as a group typically experience some cognitive difficulties (Dikmen et al., 1995). Thus, their average group scores vary across tests and cognitive domains. Brain-injured groups are, therefore, a less than ideal group for systematic comparison. Normal groups, on the other hand, perform much more

consistently across individuals and cognitive domains, and therefore can serve as a strong standard against which to compare malingerers' performances.

Because the performance of brain-injured groups in this meta-analysis varied, the malingerers' test scores were compared to both brain-injured groups (the more directly relevant comparison) and to normal controls (the more standardized comparison). Where appropriate, these comparisons were then subdivided, with one subset using only Malingering tests, and the other using all other cognitive domains. This was done on the basis of previous results indicating that using differing proportions of cognitive domains systematically affected effect sizes.

It was anticipated that the differences between malingerers and normal controls would be greater than the differences between malingerers and brain-injured participants. When the groups as a whole were compared, the difference in effect sizes between these groups was not significant, although the differences for the normal control comparison group appeared to be slightly larger than for the brain-injured comparison group. Malingerers scored on average approximately 1.2 standard deviations worse than normal controls. Again, a difference of this size means that only 38% of the malingerer and nonmalingerer distributions overlapped (Cohen, 1988).

In contrast, when the groups were divided by cognitive domain, significant differences between the comparison groups emerged. Using Malingering tests only, the difference between the malingerer/brain-injured comparison and the malingerer/normal comparison was not significant. However, using all other domains as a group, the difference between these two comparisons was significant. Examination of comparisons by individual cognitive domains revealed a complex interaction. Similar to the results using the Malingering domain only, the difference in average ES between the

comparison groups using Recognition tests was not significant. However, the results for the Recall and Visuospatial domains were significantly different between the two comparison groups. In both cases, the malingerers' performance was significantly worse (ES were significantly larger) when they were compared to the normal controls than when they were compared to brain-injured individuals. Small cell sizes prohibited contrasting the other cognitive domains across the comparison groups.

Thus, when only Malingering tests were used, differences among groups were eliminated, but when all other domains were used, significant differences emerged; this pattern of results was similar to that seen when comparing the different malingerer groups. The potential explanation for these findings presented above might also apply to the pattern between the different comparison groups. That is, it could be that Malingering tests are somehow different from tests from other cognitive domains. The Malingering tests might somehow induce all types of malingering participants to perform more poorly on these tests than on most other types of tests. Yet Malingering tests are easy enough that brain-injured participants can perform about as well as normal controls. Meanwhile, tests from other cognitive domains are possibly more difficult, particularly for brain-injured individuals. The brain-injured participants therefore do more poorly on these other domains, whereas malingerers do as poorly as they did on Malingering tests. The difference between brain-injured participants and malingerers using all other cognitive domains would be therefore smaller than either on Malingering tests or when comparing malingerers to normals. Thus, it seems that limiting analyses to the Malingering domain to try to decrease variability achieves more than that goal – it also eliminates differences between groups.

Binder and Rohling's (1996) Meta-Analysis of Financial Incentives

To summarize, the difference in test scores between experimental malingerers and normal controls ranged from $d = 1.26$ to 1.31. The aggregated differences in test performance between clinical malingerers (many of whom had suffered at least a mild head injury) and brain-injured individuals using all cognitive domains was $d = .88$. These results are noticeably different from those of Binder and Rohling (1996), who investigated the effect of financial incentives on a variety of aspects of functioning (e.g., symptom occurrence and duration, clinician ratings, return to work data) following head injury of a range of severities. The head-injured individuals in Binder and Rohling's study were compared to normal controls, a comparison that was not possible in the current meta-analysis. Across 18 studies, Binder and Rohling found that the presence of financial incentives was associated with a decrease in functioning of approximately one-half of a standard deviation relative to normal controls after head injury ($d = .47$).

It is important to note that Binder and Rohling's study differed significantly in purpose and method from the present meta-analysis. Binder and Rohling excluded experimental malingerers, while they included head-injured participants of all severities. They noted that their finding of the association between financial incentives and increased disability was stronger among participants with mild head injuries (i.e., the effect sizes were larger among mild head-injured patients, approximately $d = .89$). Binder and Rohling also did not focus exclusively on neuropsychological test scores to calculate effect sizes; they used arguably less reliable clinician and relative ratings and symptom reports.

In contrast, the studies that met inclusion criteria for the current meta-analysis that used a normal control group included only experimental malingerers. Clinical

malingerers, which are more like Binder and Rohling's head-injured group, could only be compared to brain-injured participants. Brain-injured individuals typically perform more poorly than normals on neuropsychological tests (Dikmen et al., 1995), as was demonstrated here. We would expect, then, that a clinical malingerer/brain-injured comparison would produce smaller effect sizes than a clinical malingerer/normal comparison. This hypothesis is contrary to what was found in the current meta-analysis. The test score differences produced in the clinical malingerer/brain-injured comparison in the current study (aggregated $d = .88$) is a reasonable comparison, but the difference would be expected to be smaller than that of Binder and Rohling ($d = .47$).

Why were the effect sizes so much larger in the current meta-analysis than those found by Binder and Rohling? It is likely that the different measurement methods used in the two studies accounted for the difference in effect size magnitudes. Binder and Rohling used symptom ratings, ratings by relatives and clinicians, return to work data and unidentified data on neuropsychological measures. They were not specific about the nature (i.e., validity and reliability) of the ratings used in their study. In contrast, the present meta-analysis used only data from a wide variety of neuropsychological measures. Forensic clinicians know that collateral data is essential for clarifying the veracity of a claim of injury or psychopathology. It could be that the collateral information provided by the relatives and clinicians in Binder and Rohling's study were less influenced by the litigants' exaggeration than were neuropsychological test results (which do not take into account collateral information).

When considered together, these two studies suggest that malingering or exaggeration does occur in cases involving financial incentives, and that it can impact the results of an evaluation, particularly neuropsychological test results. Therefore, it is

essential that clinicians obtain collateral information from sources that are as reliable as possible, in addition to conducting neuropsychological testing.

Limitations of the Study

As noted in the Introduction, malingering research is fraught with pitfalls that cannot be overcome in a single study. Meta-analysis combines studies with the goal that methodological differences between studies balance out. Alternatively, meta-analytic techniques can control for method variability. It could be expected, then, a meta-analysis would not have any weaknesses. However, meta-analysis remains dependent to some extent on the quality of the studies that are obtained and analyzed. If many of the studies involve a particular weakness, the studies could be excluded, leading to small cell sizes and potentially leading to a sample of studies that are not representative of the literature. Alternatively, the flawed studies could be included, but the weaknesses would have some impact on the observed results.

There was a significant degree of heterogeneity within and among the studies evaluated in this meta-analysis. This degree of heterogeneity required that the studies and samples be subdivided into smaller groupings. Many studies provided only limited information on some of these grouping variables. When there was insufficient information to permit categorization, studies had to be excluded. Likewise, some groups were excluded completely to minimize variability (e.g., comparison groups consisting of only participants with mild or severe head injury were excluded to standardize the comparison groups).

For these reasons, there was a considerable amount of attrition in the number of studies or samples available for analysis. This led to the situation that there were too

few studies using any given test to permit statistical comparison of individual cognitive tests. This is unfortunate because one of the major goals of this meta-analysis was to identify which individual tests discriminated best between malingerers and non-malingerers. The attrition also resulted in an inability to statistically compare some subgroups (e.g., Litigants). This limits the generalizability of the results to these groups. Finally, the generalizability is of these results is also limited due to the criteria for inclusion, which excluded all but published or well-known clinical neuropsychological or malingering instruments. The applicability of these results to experimental tasks like implicit memory (priming) tasks is unknown.

This study was conceptualized in terms of trying to be useful to clinicians. Its goal was to identify tests or methods of assessment that would help the clinician to identify when an individual examinee was malingering. The drawback of meta-analysis is that it is based on group data. It does not inform researchers or clinicians about how an individual malingerer might perform in an evaluation. Meta-analytic results also do not necessarily inform clinicians about ideal cutoffs for use with examinees. Thus, although the results of the current meta-analysis indicated that malingering tests and recognition tests provide the best discrimination between malingerers and normals or brain-injured individuals, it does not give a score below which examinees should be suspected of malingering. In this way, the usefulness of this meta-analysis to clinicians may be limited.

Implications of the Study

This meta-analysis demonstrated that the effect of malingering on results from neuropsychological instruments is large in both clinical and experimental settings. This

may be surprising to some psychologists in light of the smaller effect sizes found in meta-analyses of related issues [e.g., the effect of financial incentives on recovery from head injury (Binder & Rohling, 1996); the effect of mild head injury on neuropsychological test results (Binder, Rohling, & Larrabee, 1997)]. As discussed above, these studies differed from the present investigation in important ways that might account for the difference in observed effect sizes. Nevertheless, the larger results of the present meta-analysis are thought to be reasonably reflective of the actual effect sizes associated with clinical and experimental malingering.

The results of the current meta-analysis have important implications for the use of neuropsychological tests in personal injury evaluations. Unlike specific malingering tests (e.g., the Rey 15-Item Test, the Portland Digit Recognition Test), neuropsychological instruments were not designed to assess malingering. Prior to the current investigation, their validity for this purpose had not been established. The results of this study revealed that, in fact, tests from the Recognition domain performed as well as Malingering tests in discriminating between malingerers and non-malingerers. Moreover, tests of attention and psychomotor and visuospatial functioning might also provide significant discrimination between malingerers and non-malingerers if normal controls are used as the comparison group. This suggests that at least some neuropsychological instruments are valid for distinguishing between malingerers and non-malingerers.

Finally, although recognition tasks demonstrated the largest difference between malingerers and non-malingerers, tests designed strictly to assess malingering might be more appropriate for this purpose because there were no differences among malingering groups on these tests. Therefore, malingering tests might be used to identify any type of

malingerer (i.e., litigant, naïve experimental malingerer, or coached malingerers). More sensitive discrimination among groups, if necessary, could likely be achieved using traditional neuropsychological instruments.

Future research

The heterogeneity among studies revealed by this meta-analysis suggests that researchers are not addressing issues in a similar manner. This introduces an element of noise into the research that can obscure findings. Malingering researchers would do well to ensure that subject groups are more homogeneous or better defined, and that instructions to malingering groups are clearer and more similar across studies (Nies & Sweet, 1994).

The results of the Coached malingerer group in this meta-analysis were not consistent with either expectations or past research, in that the Coached group performed even more poorly on tests than the Naïve group. Lamb and colleagues' (1994) investigation of coached malingering on the MMPI-2 revealed that the type of instructions that the coached group receives has a significant impact on MMPI-2 scores. The present meta-analysis did not examine the type of instruction that the Coached subjects received. Future research might contrast the effect of information about how to elude detection and information about head injury on neuropsychological test performance. This would likely clarify our understanding of the Coached group results from the current study.

The question of which instrument(s) are best at discriminating between malingerers and non-malingerers remains unanswered. This is a clinically relevant question that was partly addressed by the present findings that tests from the recognition

and malingering domains differentiate best between the two groups. However, clinicians, and even more so, insurance adjusters, judges, and juries in litigation cases typically want a straightforward answer about which test(s) are the best. Future meta-analytic research should explore this issue in more detail.

REFERENCES

Note: Studies preceded by an asterisk were included in the meta-analysis.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

*Arnett, P.A., Hameke, T.A., & Schwartz, L. (1995). Quantitative and qualitative performance on Rey's 15-item test in neurological patients and dissimulators. *The Clinical Neuropsychologist, 9,* 17-26.

Aubrey, J.B.; Dobbs, A.R., & Rule, B.G. (1989). Laypersons' knowledge about the sequelae of minor head injury and whiplash. *Journal of Neurology, Neurosurgery, and Psychiatry, 52,* 842-846.

Baker, G.A., Hanley, J.R., Jackson, H.F., Kimmance, S., & Slade, P. (1993). Detecting the faking of amnesia: Performance differences between simulators and patients with memory impairment. *Journal of Clinical and Experimental Neuropsychology, 15,* 668-684.

Barth, J., Macciocchi, S., Giordani, B., Rimel, R., Jane, J., & Boll, T. (1983). Neuropsychological sequelae of minor head injury. *Neurosurgery, 13,* 529-533.

*Beetar, J.T., & Williams, M. (1994). Malingering response styles on the Memory Assessment Scales and the Symptom Validity Tests. *Archives of Clinical Neuropsychology, 10,* 57-72.

*Bernard, L.C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *Journal of Clinical and Experimental Neuropsychology, 12,* 715-728.

*Bernard, L.C. (1991). The detection of faked deficits on the Rey Auditory Verbal Learning Test: The effect of serial position. *Archives of Clinical Neuropsychology, 6,* 81-88.

Bernard, L.C., & Fowler, W. (1990). Assessing the validity of memory complaints: Performance of brain-damaged and normal individuals on Rey's task to detect malingering. *Journal of Clinical Psychology, 46,* 432-436.

Bernard, L.C., Houston, W., & Natoli, L. (1993). Malingering on neuropsychological memory tests: Potential objective indicators. *Journal of Clinical Psychology, 49,* 45-53.

*Bernard, L.C., McGrath, M.J., & Houston, W. (1993). Discriminating between simulated malingering and closed head injury on the Wechsler Memory Scale-Revised. *Archives of Clinical Neuropsychology, 8,* 539-551.

*Bernard, L.C., McGrath, M.J., & Houston, W. (1996). The differential effects of simulating malingering, closed head injury, and other CNS pathology on the Wisconsin Card Sorting Test: Support for the "pattern of performance hypothesis." *Archives of Clinical Neuropsychology, 11,* 231-245.

Binder, L.M. (1992). Forced-choice testing provides evidence of malingering. *Archives of Physical Medicine and Rehabilitation, 73,* 377-380.

Binder, L. M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology, 15,* 170-183.

Binder, L.M., & Rohling, M.L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *American Journal of Psychiatry, 153,* 7-10.

*Binder, L.M., & Willis, S.C. (1991). Assessment of motivation following financially compensible minor head trauma. *Psychological Assessment, 3,* 175-181.

*Binder, L.M., Villanueva, M.R., Howieson, D., & Moore, R. (1993). The Rey AVLT Recognition Memory Task measures motivational impairment after mild head trauma. *Archives of Clinical Neuropsychology, 8,* 137-147.

*Binks, P.G., Gouvier, W.D., & Waters, W.F. (1997). Malingering detection with the Dot Counting Test. *Archives of Clinical Neuropsychology, 12,* 41-46.

Bowen, M., & Littell, C. (1997). Discriminating adult normals, patients, and claimants with a pediatric test: A brief report. *The Clinical Neuropsychologist, 11,* 433-435.

Brandt, J. (1992). Detecting amnesia's impostors. In L.R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (2nd ed., pp. 156-165). New York: Guilford Press.

Brandt, J., Rubinsky, E., & Lassen, G. (1985). Uncovering malingered amnesia. *Annals of the New York Academy of Sciences, 444,* 502-503.

*Chouinard, M.-J., & Rouleau, I. (1997). The 48-Pictures Test: A two-alternative forced-choice recognition test for the detection of malingering. *Journal of the International Neuropsychological Society, 3,* 545-552.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

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

Cullum, C.M., Heaton, R.K., & Grant, I. (1991). Psychogenic factors influencing neuropsychological performance: Somatoform disorders, factitious disorders, and malingering. In H.O. Doerr & A.S. Carlin (Eds.), *Forensic neuropsychology: Legal and scientific bases* (pp. 141-171). New York: Guilford.

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

*Demakis, G.J. (1999). Serial malingering on verbal and nonverbal fluency and memory measures: An analog investigation. *Archives of Clinical Neuropsychology, 14,* 301-410.

Dikmen, S.S., Machamer, J.E., Winn, H.R., & Temkin, N.R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology, 9,* 80-90.

*Frederick, R.I., & Foster, H.G. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. *Psychological Assessment, 3,* 596-602.

Faust, D., & Ackley, M.A. (1998). Did you think it was going to be easy? Some methodological suggestions for the investigation and development of malingering detection techniques. In C.R. Reynolds (Ed.), *Detection of malingering during head injury litigation* (pp. 1-54). New York: Plenum.

Faust, D., Hart, K.J., & Guilmette, T.J. (1988). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56,* 578-582.

Faust, D., Hart, K.J., Guilmette, T.J., & Arkes, H.R. (1988). Neuropsychologists' capacity to detect adolescen malingerers. *Professional Psychology: Research and Practice, 19,* 508-515.

*Gfeller, J.D., & Cradock, M.M. (1998). Detecting feigned neuropsychological impairment with the Seashore Rhythm Test. *Journal of Clinical Psychology, 54,* 431-438.

Glass, G.V., McGaw, B., & Smith, M.L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage Publications.

Goebel, R.A. (1983). Detection of faking on the Halstead-Reitan Neuropsychological Test Battery. *Journal of Clinical Psychology, 39,* 731-742.

Gouvier, W., Prestholdt, P., & Warner, M. (1988). A survey of common misconceptions about head injury and recovery. *Archives of Clinical Neuropsychology, 3,* 331-343.

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

*Greiffenstein, M.F., Baker, W.J., & Gola, T. (1996). Motor dysfunction profiles in traumatic brain injury and postconcussion syndrome. *Journal of the International Neuropsychological Society, 2,* 477-485.

Greiffenstein, M.F. Gola, T., & Baker, W.J (1995). MMPI-2 validity scales versus domain specific measures in detection of factitious traumatic brain injury. *The Clinical Neuropsychologist, 9,* 230-240.

*Greiffenstein, M.F., Gola, T., & Baker, W.J (1996). Comparison multiple scoring methods for Rey's malingered amnesia measures. *Archives of Clinical Neuropsychology, 11,* 283-293.

*Griffin, G.A.E., Glassmire, D.M., Henderson, E.A., & McCann, C. (1997). Rey II: Redesigning the Rey Screening Test of Malingering. *Journal of Clinical Psychology, 53,* 757-766.

Grove, W.M., & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2,* 293-323.

*Guilmette, T.J., Hart, K.J., & Giuliano, A.J. (1993). Malingering detection: The use of a forced-choice method in identifying organic versus simulated memory impairment. *The Clinical Neuropsychologist, 7,* 59-69.

Guilmette, T.J., Hart, K.J., Giuliano, A.J., & Leininger, B.E. (1994). Detecting simulated memory impairment: Comparison of the Rey Fifteen-Item test and the Hiscock Forced-Choice Procedure. *The Clinical Neuropsychologist, 8,* 283-294.

*Guilmette, T.J., Whelihan, W., Sparadeo, F.R., & Buongiorno, G. (1994). Validity of neuropsychological test results in disability evaluations. *Perceptual and Motor Skills, 78,* 1179-1186.

*Guilmette, T.J., Whelihan, W., Hart, K.J., Sparadeo, F.R., & Buongiorno, G. (1996). Order effects in the administration of a forced-choice procedure for detection of malingering in disability claimants' evaluations. *Perceptual and Motor Skills, 83,* 1007-1016.

Hall, H.V., Shooter, E.A., Craine, J., & Paulsen, S. (1991). Explicit alternative testing: A trilogy of studies on faked memory deficits. *Forensic Reports, 4,* 259-279.

Hays, J., Emmons, J., & Lawson, K. (1993). Psychiatric norms for the Rey 15-item visual memory test. *Perceptual and Motor Skills, 76,* 1331-1334.

Hayward, L., Hall, W., Hunt, M., & Zubrick, S.R. (1987). Can localised brain impairment be simulated on neuropsychological test profiles? *Australian and New Zealand Journal of Psychiatry, 21,* 87-93.

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

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

*Hiscock, C.K., Branham, J.D., & Hiscock, M. (1994). Detection of feigned cognitive impairment: The two-alternative forced-choice method compared with selected conventional tests. *Journal of Psychophysiology and Behavioral Assessment, 16,* 95-110.

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.

Horton, K.D., Smith, S.A., Barghout, N.K., & Connolly, D. A. (1992). The use of indirect memory tests to assess malingered amnesia: A study of metamemory. *Journal of Experimental Psychology: General, 121,* 326-351.

*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,* 128-138.

Iverson, G.L. (1995). Qualitative aspects of malingered memory deficits. *Brain Injury, 9,* 35-40.

*Iverson, G.L., & Franzen, M.D. (1994). The Recognition Memory Test, Digit Span, and Knox Cube Test as markers of malingered memory impairment. *Assessment, 1,* 323-334.

*Iverson, G.L., & Franzen, M.D. (1996). Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology, 18,* 38-51.

*Iverson, G.L., & Franzen, M.D. (1998). Detecting malingered memory deficits with the Recognition Memory Test. *Brain Injury, 12,* 275-282.

*Iverson, G.L., Franzen, M.D., & McCracken, L.M. (1991). Evaluation of an objective assessment technique for the detection of malingered memory deficits. *Law and Human Behavior, 15,* 667-676.

*Iverson, G.L., Franzen, M.D., & McCracken, L.M. (1994). Application of a forced-choice memory procedure designed to detect experimental malingering. *Archives of Clinical Neuropsychology, 9,* 437-450.

Johnson, B.T. (1989). *DSTAT: Software for the meta-analytic review of research literatures.* Hillsdale, NJ: Erlbaum.

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

*King, J.H., Gfeller, J.D., & Davis, H.P. (1998). Detecting simulated memory impairment with the Rey Auditory Verbal Learning Test: Implications of base rates and study generalizability. *Journal of Clinical and Experimental Neuropsychology, 20,* 603-612.

*Klimczak, N.J., Donovick, P.J., & Burright, R. (1997). The malingering of multiple sclerosis and mild traumatic brain injury. *Brain Injury, 11,* 343-352.

Klonoff, P.S., & Lamb, D.G (1998). Mild head injury, significant impairment on neuropsychological test scores, and psychiatric disability. *The Clinical Neuropsychologist, 12,* 31-42.

Lamb, D.G., Berry, D.T.R., Wetter, M.W., & Baer, R.A. (1994). Effects of two types of information on malingering of closed head injury on the MMPI-2: An analog investigation. *Psychological Assessment, 6,* 8-13.

Larrabee, G.J., & Curtiss, G. (1995). Construct validity of various verbal and visual memory tests. *Journal of Clinical and Experimental Neuropsychology, 17,* 536-547.

Larrabee, G.J., Kane, R.L., Schuck, J.R., & Francis, D.J. (1985). Construct validity of various memory testing procedures. *Journal of Clinical and Experimental Neuropsychology, 7,* 239-250.

*Lee, G., Loring, D., & Martin, R. (1992). Rey's 15-item visual memory test for the detection of malingering: Normative observations on patients with neurological disorders. *Psychological Assessment, 4,* 43-46.

Lees-Haley, P.R., Smith, H.H., Williams, C.W., & Dunn, J.T. (1996). Forensic neuropsychological test usage: An empirical study. *Archives of Clinical Neuropsychology, 2,* 45-51.

Leininger, B.E., Gramling, S.E., Farrell, A.D., Kreutzer, J.S., & Peck, E.A. (1990). Neuropsychological deficits in symptomatic minor head injury patients after concussion and mild concussion. *Journal of Neurology, Neurosurgery, and Psychiatry.*

Leonberger, F.T., Nicks, S.D., Larrabee, G.J., & Goldfader, P.R. (1992). Factor structure of the Wechsler Memory Scale-Revised within a comprehensive neuropsychological battery. *Neuropsychology, 6,* 239-249.

Levin, H.S. (1982). Pathophysiologic mechanisms. In H.S. Levin, A.L. Benton, & Grossman, R.G. (Eds.), *Neurobehavioral consequences of closed head injury* (pp. 3-45). New York: Oxford University Press.

Levin, H.S., Lilly, M.A., Papanicoulaou, A., & Eisenberg, H.M. (1992). Posttraumatic and retrograde amnesia after closed head injury. In L.R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (2nd ed.) (pp. 290-308). New York, NY: Guilford.

Lezak, M. (1995). *Neuropsychological assessment* (3rd edition). New York: Oxford University Press.

Light, R.J., & Pillemer, D.B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard 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.

*Martin, R.C., Franzen, M.D., & Orey, S. (1998). Magnitude of error as a strategy to detect feigned memory impairment. *The Clinical Neuropsychologist, 12,* 84-91.

*Martin, R.C., Gouvier, W.D., Todd, M.E., Bolter, J.F., & Niccolls, R. (1993). Effects of task instruction on malingered memory performance. *Forensic Reports, 5,* 393-397.

Martin, R.C., Hayes, J.S., & Gouvier, W.D. (1996). Differential vulnerability between postconcussion self-report and objective malingering tests in identifying simulated mild head injury. *Journal of Clinical Neuropsychology, 18,* 265-275.

*Mensch, A.J. & Woods, D.J. (1986). Patterns of feigning brain damage on the LNNB. *International Journal of Clinical Neuropsychology, 8,* 59-63.

Meyers, J.E., & Volbrecht, M. (1998). Validtation of Reliable Digits for the detection of malingering. *Assessment, 5,* 303-307.

Milanovich, J.R., Axelrod, B.N., & Millis, S.R. (1996). Validation of the Simulation Index-Revised with a mixed clinical population. *Archives of Clinical Neuropsychology, 11,* 53-59.

Miller, H., & Cartlidge, N. (1972). Simulation and malingering after injuries to the brain and spinal cord. *Lancet, 1,* 580-586.

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

Millis, S.R. (1994). Assessment of motivation and memory with the Recognition Memory Test after financially compensible mild head injury. *Journal of Clinical Psychology, 50,* 601-605.

Millis, S.R., & Kler, S. (1995). Limitations of the Rey Fifteen-Item Test in the detection of malingering. *The Clinical Neuropsychologist, 9,* 241-244.

Millis, S.R., & Putnam, S.H. (1994). The Recognition Memory Test in the assessment of memory impairment after financially compensible mild head injury: A replication. *Perceptual and Motor Skills, 79,* 384-386.

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,* 463-471.

*Mittenberg, W., Azrin, R., Millsaps, C., & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale-Revised. *Psychological Assessment, 5,* 34-40.

Mittenberg, W., DiGiulio, D.V., Perrin, S., & Bass, A.E. (1989). Symptoms following mild head injury: Expectations as etiology. *The Clinical Neuropsychologist, 3,* 297.

*Mittenberg, W., Rotholc, A., Russell, E. & Heilbronner, R. (1996). Identification of malingered head injury on the Halstead-Reitan Battery. *Archives of Clinical Neuropsychology, 11,* 271-281.

*Mittenberg, W., Theroux-Fichera, S., Zielinski, R., & Heilbronner, R.L. (1995). Identification of malingered head injury on the Wechsler Adult Intelligence Scale-Revised. *Professional Psychology: Research and Practice, 26,* 491-498.

Nies, K.J., & Sweet, J.J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology, 9,* 501-552.

*Osimani, A., Alon, A., Berger, A., & Abarbanel, J.M. (1997). Use of the Stroop phenomenon as a diagnostic tool for malingering. *Journal of Neurology, Neurosurgery, and Psychiatry, 62,* 617-621.

Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology, 47,* 409-410.

Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills, 57,* 367-372.

Pankratz, L., & Binder, L.M. (1997). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 223-236). New York: Guilford.

Prigatano, G.P., & Amin, K. (1993). Digit Memory Test: Unequivocal cerebral dysfunction and suspected malingering. *Journal of Clinical and Experimental Neuropsychology, 15,* 537-546.

*Prigatano, G.P., Smason, I., Lamb, D.G., & Bortz, J.J. (1997). Suspected malingering and the Digit Memory Test: A replication and extension. *Archives of Clinical Neuropsychology, 12,* 609-619.

*Rapport, L.J., Farchione, T.J., Coleman, R.D., & Axelrod, B.N. (1998). Effects of coaching on malngered motor function profiles. *Journal of Clinical and Experimental Neuropsychology, 20,* 89-97.

*Rees, L.M., Tombaugh, T.N., Gansler, D.A., & Moczynski, N.P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment, 10,* 10-20.

Reitan, R.M., & Wolfson, D. (1996). The question of validity of neuropsychological tests scores among head-injured Litigants: Development of a dissimulation index. *Archives of Clinical Neuropsychology, 11,* 573-580.

Reitan, R.M., & Wolfson, D. (1997). Consistency of neuropsychological test scores of head-injured subjects involved in litigation compared with head-injured subjects not involved in litigation: Development of the Retest Consistency Index. *The Clinical Neuropsychologist, 11,* 69-76.

Reynolds, C.R. (1998). *Detection of malingering during head injury litigation.* New York: Plenum.

Rimel, R., Giordani, B., Barth, J., Boll, T., & Jane, J. (1981). Disability caused by minor head trauma. *Neurosurgery, 9,* 221-228.

Rogers, R. (1997). Current status of clinical methods. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 373-397). New York: Guilford.

Rogers, R., Bagby, R.M., & Chakraborty, D. (1993). Feigning schizophrenic disorders on the MMPI-2: Detection of coached simulators. *Journal of Personality Assessment, 60,* 215-226.

Rogers, R., & Cavanaugh, J.L. (1983). "Nothing but the truth": A reexamination of malingering. *Journal of Law and Psychiatry,11,* 443-460.

Rogers, R., Harrell, E.H., & Liff, C.D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review, 13,* 255-274.

*Rose, F.E., Hall, S., & Szalda-Petree, A.D. (1995). Portland Digit Recognition Test-Computerized: Measuring response latency improves the detection of malingering. *The Clinical Neuropsychologist, 9,* 124-134.

*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.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118,* 183-192.

Schacter, D., & Crovitz, H. (1977). Memory function after closed head injury: A review of the quantitative research. *Cortex, 13,* 105-176.

*Schagen, S., Schmand, B., deSterke, S., & Lindeboom, J. (1997). Amsterdam Short Term Memory Test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical and Experimental Neuropsychology, 19,* 43-51.

*Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T., & Hamburger, J.L. (1998). Cognitive complaints in patients after whiplash injury: The impact of malingering. *Journal of Neurology, Neurosurgery, and Psychiatry, 64,* 339-342.

Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1,* 115-129.

Schretlen, D., Brandt, J., Krafft, L. & VanGorp, W. (1991). Some caveats in using the Rey 15-item memory test to detect malingered amnesia. *Psychological Assessment, 3,* 667-672.

Schwartz, S.M., Gramling, S.E., Kerr, K.L., & Morin, C. (1998). Evaluation of intellect an ddeficit-specific information on the ability to fake memory deficits. *International Journal of Law and Psychiatry, 21,* 261-272.

Simon, M. (1994). The use of the Rey memory test to assess malingering in criminal defendants. *Journal of Clinical Psychology, 50,* 913-917.

Sinnett, E. R., & Holen, M.C. (1995). Chance performance in the absence of norms for neuropsychological tests: A clinical note. *Journal of Clinical Psychology, 51,* 400-402.

Slick, D., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's Forced-Choice Memory Test. *Journal of Clinical and Experimental Neuropsychology, 16,* 472-481.

*Slick, D.J., Hopp, G., Strauss, E., & Spellacy, F.J. (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.

*Spanos, N.P., Radtke, H.L., Bertrand, L.D., Addie, D.L., & Drummond, J. (1982). Disoganized recall, hypnotic amnesia, and subjects' faking: More disconfirmatory evidence. *Psychological Reports, 50,* 383-389.

Spreen, O., & Strauss, E. (1998). Malingering and symptom validity testing. In O. Spreen & E. Strauss, *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd edition) (pp. 666-685). New York: Oxford University Press.

Strauss, E., Spellacy, F., Hunter, M., & Berry, T. (1994). Assessing believable deficits on measures of attention and information processing capacity. *Archives of Clinical Neuropsychology, 9,* 483-490.

Suhr, J., Tranel, D., Wefel, J., & Barrash, J. (1997). Memory impairment after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. *Journal of Clinical and Experimental Neuropsychology, 19,* 500-514.

*Tenhula, W.N., & Sweet, J.J. (1996). Double cross-validation of the Booklet Category Test in detecting malingered traumatic brain injury. *The Clinical Neuropsychologist, 10,* 104-116.

Thornton, A.E., & Raz, N. (1997). Memory impairment in multiple sclerosis: A quantitative review. *Neuropsychology, 11,* 357-366.

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

Trueblood, W., & Binder, L.M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers. *Archives of Clinical Neuropsychology, 12,* 13-27.

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,* 578-590.

*Tsushima, W.T., & Wong, J.M. (1992). Comparison of legal and medical referrals to neuropsychological examination following head injury. *Forensic Reports, 5,* 359-366.

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

Wetter, M.W., Baer, R.A., Berry, D.T.R., & Reynolds, S.K. (1994). The effect of symptom information on faking on the MMPI-2. *Assessment, 1,* 199-207.

Wetter, M.W., & Deitsch, S.E. (1996). Faking specific disorders and temporal response consistency on the MMPI-2. *Psychological Assessment, 8,* 39-47.

Wiggins, E.C., & Brandt, J. (1988). The detection of malingered amnesia. *Law and Human Behavior, 12,* 57-78.

Williams, J.M. (1998). The malingering of memory disorder. In C.R. Reynolds (Ed.), *Detection of malingering during head injury litigation* (pp. 105-132). New York: Plenum.

Williams, R.W., & Carlin, M. (1999). Malingering on the WAIS-R among disability claimants and applicants for vocational assistance. *American Journal of Forensic Psychology, 17,* 35-45.

*Wogar, M.A., Van den Broek, M.D., Bradshaw, C.M., & Szabadi, E. (1998). A new performance-curve method for the detection of simulated cognitive impairment. *British Journal of Clinical Psychology, 37,* 327-339.

Wong, J.L., Regennitter, R.P., & Barrios, F. (1994). Base rate and simulated symptoms of mild head injury among normals. *Archives of Clinical Neuropsychology, 9,* 411-425.

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.

Zielinski, J.J. (1994). Malingering and defensiveness in the neuropsychological assessment of mild traumatic brain injury. *Clinical Psychology: Science and Practice, 1,* 169-184.

Appendix A:

Coding Sheet

EFFECT SIZE DATA SHEET **STUDY #**
Authors: **Journal:** **Year:**

Subject groups: **Setting:**

Variables controlled for: **Indexes used (tests given):**

Group 1 (Malingerers):

N:	Age(SD):	Educ(SD):	%Male:
Sample:		Instructions:	
Compensation:		Coach/Naive:	
Exclusion criteria:			

Group 2 (Patients):

N:	Age(SD):	Educ(SD):	%Male:
In/Outpt.:	Dx:	Criteria for injury:	
Time since injury:			
Exclusion criteria:			

Group 3 (Controls):

N:	Age(SD):	Educ(SD):	%Male:
In/Outpt.:	Dx:	Criteria for injury:	
Time since injury:			
Sample:	Instructions:		
Compensation:	Coach/Naive:		
Exclusion criteria:			

Test:		ES: d=	g=
Mal grp:	n:	X=	SD=
Pt/Control:	n:	X=	SD=
t=	F=	other=	

Test:		ES: d=	g=
Mal grp:	n:	X=	SD=
Pt/Control:	n:	X=	SD=
t=	F=	other=	

Appendix B:

Neuropsychological Instruments and Scores

Instrument and Scores	Cognitive Domain
Rey 15-Item test Number of items recalled; Number of items correctly located; Number of rows correctly recalled; Number of rows correctly sequenced	Malingering
21-Item Test Free recall; Forced choice	Malingering
48 Pictures Test Total score; Immediate recognition; Delayed recognition	Malingering
Benton Visual Retention Test (BVRT) Number Correct; Errors	Visuospatial
Category Test Number of errors	Executive
Hiscock Forced Choice Procedure and Abbreviation/ Digit Memory Test (DMT/HFCP) Delays: 2, 7 and 15 seconds or 5, 10, and 15 seconds	Malingering
Dot Counting (Lezak) Grouped and ungrouped dots	Malingering
Finger Agnosia	Sensory-Perceptual
Fingertip Writing	Sensory-Perceptual
Grip Strength Dominant and nondominant hands	Psychomotor
Letter Memory Test Percent correct	Malingering
Luria-Nebraska Neuropsychological Battery	
Arithmetic	Attention
Expressive	Language
Intellectual	Intellectual
Memory	Recall
Motor	Psychomotor
Reading	Language
Receptive	Language
Rhythm	Psychomotor
Tactile	Psychomotor
Visual	Visuospatial
Writing	Language
Multi-Digit Modality Test (MDMT) Delays: 2, 7 and 15 seconds or 5, 10, and 15 seconds	Malingering
Paced Auditory Serial Addition Test (PASAT) Total correct	Attention
Portland Digit Recognition Test (PDRT) and PDRT-27 Delays: 2, 7 and 15 seconds or 5, 10, and 15 seconds	Malingering
Grooved Pegboard Test Dominant and nondominant hands	Psychomotor
Rey Auditory Verbal Learning Test; California Verbal Learning Test (RAVLT/CVLT) Trial 1; Total of trials 1 to 5; Distractor (List B); Delay (Trial 6); Total recall; Total recognition; Recognition true and false positives; Learning curve	Recall or Recognition (as appropriate)
Recognition Memory Test (RMT) Words recall; Faces recall	Recognition
Rey Recognition Word List	Recognition

Instrument and Scores	Cognitive Domain
Rey-Osterreith Complex Figure Test	Visuospatial (copy)
Copy; Recall	Recall (recall)
Seashore Rhythm Test	Attention
Number correct	
Speech Sounds Perception Test	Attention
Number of errors	
Stroop Test	Attention
Finger Tapping Test	Psychomotor
Dominant and nondominant hands	
Test of Memory Malingering (TOMM)	Malingering
Trial 1; Trial 2; Retention correct and incorrect; Latency for all scores	
Test of Nonverbal Intelligence (TONI) forced choice	Malingering
Forced choice # correct; Slope; Consistency ratio; Slope x Consistency ratio	
Tactual Performance Test	Psychomotor (time per block,
Total time per block; memory; location	location); Recall (memory)
Trails A & B	Executive
Time to completion; number correct/min.	
Victoria Symptom Validity Test (VSVT)	Malingering
Easy item recall and response time; Hard item recall and response time	
WAIS-R Performance IQ[1]	Intellectual
WAIS-R Verbal IQ[1]	Intellectual
Information	Intellectual
Arithmetic	Attention
Vocabulary	Language
Digit Span – Raw scores and Scaled scores	Attention
Digit Symbol	Psychomotor
Picture Completion	Visuospatial
Picture Arrangement	Visuospatial
Object Assembly	Visuospatial
Block Design	Visuospatial
Wisconsin Card Sorting Test (WCST)	Executive
Number of categories	
Wechsler Memory Scale-Revised (WMS-R)[2]	
Figural Memory	Recall
Logical Memory I	Recall
Logical Memory II	Recall
Mental Control	Attention
Verbal Paired Associates	Recall
Visual Paired Associates	Recall
Visual Reproduction I	Visuospatial
Visual Reproduction II	Recall
Visual Span	Visuospatial

Note: [1] If a study gave both the FSIQ, VIQ or PIQ and individual subtest scores, only the subtest scores were used to avoid non-independence of data and to avoid loss of data through using IQ summary scores.
[2] WMS-R Index scores were not used to avoid non-independence of data and to avoid loss of data through using summary scores.

Appendix C:

Stem-and Leaf Plot for Brain-Injured and Normal Controls

Brain-Injured	Normal Controls
332211.0.	
9999777765.0.7788	
	20.1.233344
9987655.1.557889	
	40.2.3344
	75.2.99

Appendix D:

Studies Included and Excluded from Analysis

Table D.1

Studies Excluded from Analysis

Author	Reason for Exclusion
Benton & Spreen (1961)	ES could not be estimated
Bernard & Fowler (1995)	No information re: Comparison group head injury severity
Bernard et al. (1993)	Same (but smaller) sample as Bernard (1990; 1991)
Bickart et al. (1991)	ES could not be estimated
Binder (1993)	Same (but smaller) sample as Binder & Willis (1991)
Binder & Kelly (1996)	ES could not be estimated
Boone et al. (1995)	ES could not be estimated
Brandt et al. (1985)	ES could not be estimated
Bruhn & Reed (1975)	Data not presented usably
Cochrane et al. (1998)	Data not presented usably
Fox et al. (1995)	ES could not be estimated
Franzen & Martin (1996)	No control group, ES could not be estimated
Frederick et al. (1994)	ES could not be estimated
Gasquoine (1997)	ES could not be estimated
Gass & Russell (1991)	No control group
Goebel (1983)	ES could not be estimated
Greiffenstein et al. (1994)	Comparison group in litigation
Gudjonsson & Shackleton (1986)	Data not presented usably
Hall et al. (1991)	Data not presented usably; missing data
Hayward et al. (1987)	ES could not be estimated
Johnson et al. (1998)	Data not presented usably
Karzmark et al. (1995)	No control group, ES could not be estimated
Lees-Haley (1991)	No information re: Comparison group head injury severity
Lees-Haley (1997)	No control group
Lees-Haley & Fox (1990)	No control group
Lees-Haley et al. (1991)	No information re: Comparison group head injury severity
Leininger et al. (1990)	Comparison group included only mild head injury
Martin et al. (1996)	Comparison group included only mild head injury
McKinzey & Russell (1997)	ES could not be estimated
McKinzey & Russell (1997)	Data not presented usably
Meyers & Volbrecht (1998)	Comparison group included only mild head injury
Millis (1992)	Comparison group included only mild head injury
Millis (1994)	Comparison group included only mild head injury
Millis & Kler (1995)	Comparison group included only severe head injury
Millis & Putnam (1994)	Comparison group included only severe head injury
Millis et al. (1995)	Comparison group included only severe head injury

Author	Reason for Exclusion
Millis et al. (1998)	Instrument could not be categorized into cognitive domains
Ridenour et al. (1996)	ES could not be estimated
Ridenour et al. (1998)	ES could not be estimated
Schretlen et al. (1991)	ES could not be estimated
Slick et al. (1994)	Same (but smaller) sample as Slick et al. (1996)
Snibbe et al. (1980)	No control group
Strauss et al. (1994)	Comparison group included only mild head injury
Suhr et al. (1997)	Comparison group included mild and severe head injury
Taylor et al. (1996)	Malingering group included only whiplash patients (no documented head injury)
Trueblood & Schmidt (1993)	No control group
Trueblood (1994)	Same sample as Trueblood & Schmidt (1993)
Wiggins & Brandt (1988)	ES could not be estimated
Williams & Carlin (1999)	No information re: Comparison group head injury severity

Table D.2

Samples Excluded Due to Outlying Effects

Author	Reason for Exclusion (*d*)
Iverson & Franzen (1998)	4.62
Rees et al. (1998)	5.31; 5.10
Schagen et al. (1997)	4.24
Tombaugh (1997)	5.43; 5.18; 4.32

Table D.3

Effect Sizes Calculated for Each Comparison Listed by Study

Author	*d*	G1 *n*	G2 *n*
Arnett et al. (1995) a	.77	49	34
Arnett et al. (1995) b	.91	25	25
Beetar & Williams (1994)	1.17	30	30
Bernard (1990)	.82	30	28
Bernard (1991)	.44	89	44
Bernard, McGrath, & Houston (1993)	1.36	24	80
Bernard et al. (1996)	.95	18	38
Binder et al. (1993)	.31	17	15
Binder & Willis (1991)	1.83	13	38
Binks et al. (1997)	1.53	27	27
Chouinard & Rouleau (1997)	1.81	27	27
Coleman et al. (1998)	.19	43	44
Demakis (1999)	.69	84	14
Frederick & Foster (1991)	1.25	58	57
Gfeller & Cradock (1998)	1.56	56	57
Greiffenstein, Baker & Gola (1996)	.12	66	29
Greiffenstein, Gola & Baker (1996)	.74	90	60
Griffin et al. (1997)	1.21	20	20
Guilmette et al. (1993)	.73	68	56
Guilmette et al. (1994)	1.79	65	65
Guilmette et al. (1996)	2.21	29	20
Heaton et al. (1978)	.39	20	20
Hiscock et al. (1994) a	.13	16	16
Hiscock et al. (1994) b	.75	71	71
Inman et al. (1998) a	.82	71	71
Inman et al. (1998) b	2.18	42	32
Inman et al. (1998) c	1.66	25	28
Iverson & Franzen (1994)	1.70	30	30
Iverson & Franzen (1996)	2.21	20	20
Iverson & Franzen (1998)	4.10	20	20
Iverson et al. (1991)	1.43	21	20
Iverson et al. (1994)	1.67	20	20
Johnson & Lesniak-Karpiak (1997)	.34	29	29
King et al. (1998) a	1.31	50	50
King et al. (1998) b	1.48	20	40
Klimczak et al. (1997)	1.37	27	14
Lee et al. (1992)	1.05	16	70
Martin et al. (1992)	1.40	38	28
Martin et al. (1993) a	2.22	20	16
Martin et al. (1993) b	.70	17	16
Martin et al. (1998) a	.59	15	30

Author	*d*	G1 *n*	G2 *n*
Martin et al. (1998) b	.22	7	30
Mensch & Woods (1986)	2.12	32	32
Mittenberg et al. (1993)	.39	39	39
Mittenberg et al. (1995)	.23	80	80
Mittenberg et al. (1996)	.10	67	67
Osimani et al. (1997)	.44	10	10
Prigatano et al. (1997)	2.07	6	7
Rapport et al. (1998)	1.29	32	30
Rees et al. (1998) a	1.31	32	29
Rees et al. (1998) b	2.43	20	20
Rees et al. (1998) c	2.40	25	19
Rees et al. (1998) d	2.08	13	13
Rose et al. (1995)	1.37	20	20
Rose et al. (1998)	.38	29	28
Schagen et al. (1997)	1.27	29	28
Schmand et al. (1998)	.79	10	20
Slick et al. (1996)	.13	20	31
Spanos et al. (1982)	1.26	43	32
Tenhula & Sweet (1996) a	3.85	20	20
Tenhula & Sweet (1996) b	.94	28	26
Tombaugh (1997) a	1.75	17	24
Tombaugh (1997) b	3.72	20	28
Tombaugh (1997) c	.11	36	103
Tsushima & Wong (1992)	1.01	34	32
Wogar et al. (1998)	.18	20	25

Note: G1 *n* = number of participants in malingering group.
G2 *n* = number of participants in comparison group.

Appendix E:

Study Characteristics for Study Quality Comparison

Study Characteristics	All Studies	Good
Number of Studies	29	24
Malingering Group		
Age	29.09 (6.82)	28.23 (6.40)
Education	13.87 (1.38)	12.96 (1.32)
Percent male	58% (27%)	55% (27%)
Head injury severity	11% mild 2% moderate 4% range 83% n/a	4% mild 2% range 93% n/a
Comparison Group		
Age	33.96 (4.70)	33.34 (3.27)
Education	12.90 (.78)	12.96 (.76)
Percent male	56% (17%)	55% (17%)
Time since injury (months)	44.18 (28.19)	43.84 (29.5)

Note: Values in parentheses represent standard deviations. n/a = not applicable.